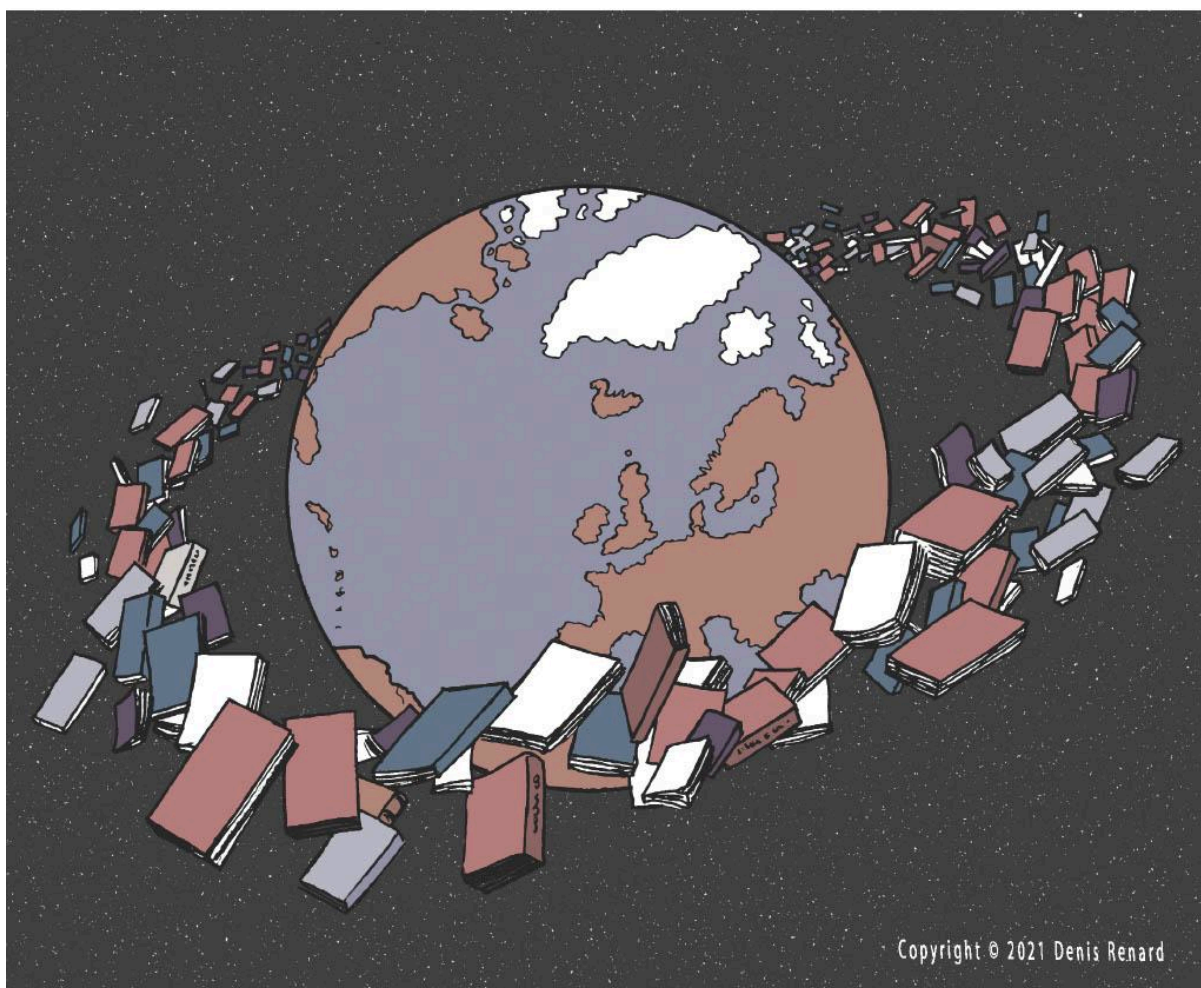


# The web, from English to multilingual

by [Marie Lebert](#)



This ebook includes two parts:

1. "The web, from English to multilingual" (December 2019)
2. "Languages on the web - a timeline" (May 2025)

May 2025

Community Texts of the Internet Archive

Licence CC BY-NC-SA 4.0

# The web, from English to multilingual



By Marie Lebert, 9 December 2019.

## THE BOOK IN THE COMMUNITY TEXTS OF THE INTERNET ARCHIVE

After the invention of the web in 1990, internet users with a native language other than English reached 5 percent in 1994, 20 percent in 1998, 50 percent in 2000 and 75 percent in 2015. Many people helped promote their own language and culture and other languages and cultures — sometimes on their free time and often using English as a lingua franca — for the web to become truly multilingual. This book based on many interviews is a tribute to their hard work and dedication.

---

[\[French version\]](#)

[\[Spanish version\]](#)

- 
- \* [Overview](#)
  - \* [About the internet](#)
  - \* [About encoding](#)
  - \* [About internationalisation](#)
  - \* [About multilingualism](#)
  - \* [About website localisation](#)
  - \* [About remote collaboration](#)
  - \* [About e-texts](#)
  - \* [About e-books](#)
  - \* [About the press](#)
  - \* [About a shift in jobs](#)
  - \* [About copyright](#)
  - \* [About creative copyright](#)
  - \* [About bookstores](#)
  - \* [About e-bookstores](#)
  - \* [About authors](#)
  - \* [About libraries](#)
  - \* [About librarians](#)
  - \* [About digital libraries](#)
  - \* [About treasures of the past](#)
  - \* [About library catalogues](#)
  - \* [About linguistic resources](#)

- \* [About dictionaries](#)
  - \* [About encyclopedias](#)
  - \* [About journals](#)
  - \* [About resources for teaching](#)
  - \* [About resources for translators](#)
  - \* [About terminology databases](#)
  - \* [About machine translation](#)
  - \* [About computer-assisted translation](#)
  - \* [About free machine translation services](#)
  - \* [About the Ethnologue](#)
  - \* [About minority languages](#)
  - \* [About endangered languages](#)
  - \* [Timeline](#)
- 

## Overview

After the invention of the web in 1990, internet users with a native language other than English reached 5 percent in 1994, 20 percent in 1998, 50 percent in 2000 and 75 percent in 2015.

Brian King, director of the Worldwide Language Institute (WWLI) in Europe, brought up the concept of “linguistic democracy” in September 1998 in an email interview: “Whereas ‘mother-tongue education’ was deemed a human right for every child in the world by a UNESCO report in the early 1950s, ‘mother-tongue surfing’ may very well be the Information Age equivalent. If the internet is to truly become the Global Network that it is promoted as being, then all users, regardless of language background, should have access to it. To keep the internet as the preserve of those who, by historical accident, practical necessity, or political privilege, happen to know English, is unfair to those who don’t.”

Maria Victoria Marinetti, a Mexican engineer who was working as a Spanish-language teacher and translator abroad, wrote in August 1999: “It is very important to be able to communicate in various languages. I would even say this is mandatory, because the information given on the internet is meant for the whole world, so why wouldn’t we get this information in our language or in the language we wish? Worldwide information, but no broad choice for languages, this would be quite a contradiction, wouldn’t it?”

Internet users living outside the United States reached 50 per cent in summer 1999. Jean-Pierre Cloutier, editor of *Chroniques de Cybérie*, a weekly French-language online report of internet news, wrote in August 1999: “We passed a milestone this summer. Now more than half of the users of the internet live outside the United States. Next year, more than half of all users will have a native language other than English, compared with only 5 percent five years ago. Isn’t that great?”

The number of internet users with a native language other than English did reach 50 percent in summer 2000, 52.5 percent in summer 2001, 57 percent in December 2001, 59.8 percent in April 2002 and 64.4 percent in September 2003, according to the marketing consultancy Global Reach.

Fifteen years after the invention of the web, the monthly Wired stated in August 2005 that “less than half of the web is commercial, and the other half is run by passion.” According to the French daily Le Monde on 19 August 2005, “the three powers of the internet — ubiquity, variety and interactivity — make its potential use quasi infinite.”

Many people helped promote their own language and culture and other languages and cultures — sometimes on their free time and often using English as a lingua franca — for the web to become truly multilingual. This book based on many interviews is a tribute to their hard work and their dedication.

---

## About the internet

Eduard Hovy, head of the Natural Language Group at USC/ISI (University of Southern California / Information Sciences Institute), wrote in August 1998 in an email interview: “The internet is, as I see it, a fantastic gift to humanity. It is, as one of my graduate students recently said, the next step in the evolution of information access. A long time ago, information was transmitted orally only; you had to be face-to-face with the speaker. With the invention of writing, the time barrier broke down — you can still read Seneca and Moses. With the invention of the printing press, the access barrier was overcome — now anyone with money to buy a book can read Seneca and Moses. And today, information access becomes almost instantaneous, globally; you can read Seneca and Moses from your computer, without even knowing who they are or how to find out what they wrote; simply open AltaVista and search for ‘Seneca’. This is a phenomenal leap in the development of connections between people and cultures.”

Henri “Henk” Slettenhaar had a long career as a communication systems specialist in Geneva, Switzerland, and in Silicon Valley, California. He was fluent in three languages. He spent his childhood in Holland, taught his courses in English, and learned French while living in neighbouring France. He joined the CERN (European Organisation for Nuclear Research) in Geneva in 1958 to work on the first digital computer and the first digital networks. He joined SLAC (Stanford Linear Accelerator Center) to build a film digitiser in 1966 and a monitoring system in 1983. He settled back in Geneva to teach communication technology at Webster University for 25 years, and ran its Telecom Management Program from 2000.

He wrote in December 1998: “I can’t imagine my professional life without the internet. Most of my communication is now via email. I have been using email for the last twenty years, most of that time to keep in touch with colleagues in a very narrow field. Since the explosion of the internet, and especially the invention of the web, I communicate mainly by email. Most of my presentations are now on the web and the courses I teach are all web-extended. All the details of my Silicon Valley tours are on the web. Without the internet we wouldn’t be able to function. And I use the internet as a giant database. I can find information today with the click of a mouse.”

“I also see multilingualism as a very important issue. Local communities that are on the web should principally use the local language for their information. If they want to present it to the world community as well, it should be in English too. I see a real need for bilingual websites. I am

delighted there are so many offerings in the original language now. I much prefer to read the original with difficulty than getting a bad translation.”

He added in August 1999: “There are two main categories of websites in my opinion. The first one is the global outreach for business and information. Here the language is definitely English first, with local versions where appropriate. The second one is local information of all kinds in the most remote places. If the information is meant for people of an ethnic and/or language group, it should be in that language first, with perhaps a summary in English.”

He wrote the next year about experiencing “the explosion of mobile technology. The mobile phone has become for many people, including me, the personal communicator which allows you to be anywhere at anytime and still be reachable. But the mobile internet is still a dream. The new services on mobile (GSM) phones are extremely primitive and expensive (WAP = Wait and Pay). Multilingualism has expanded greatly. Many e-commerce websites are multilingual now, and there are companies that sell products which make localisation possible.”

He wrote again in July 2001: “I am experiencing a tremendous change with having a ‘broadband’ connection at home. To be connected at all times is so completely different from dial-up. I now receive email as soon as it arrives, I can listen to my favourite radio stations wherever they are. I can listen to the news when I want to. Get the music I like all the time. The only thing which is missing is good quality real-time video. The bandwidth is too low for that. I now have a wired and a wireless LAN [Local Area Network] in my home. I can use my laptop anywhere in the house and outside, even at my neighbours’ house and still being connected. With the same technology I am now able to use my wireless LAN card in my computer when I travel. For instance, during my recent visit to Stockholm, there was connectivity in the hotel, the conference center, the airport, and even in the Irish pub!”

Pierre Schweitzer, designer of the mobile device @folio, wrote in December 2006: “The luck we all have is to live here and now this fantastic change. When I was born in 1963, computers didn’t have much memory. Today, my music player could hold billions of pages, a true local library. Tomorrow, by the combined effect of the Moore Law and the ubiquity of networks, we will have instant access to works and knowledge. We won’t be much interested any more on which device to store information. We will be interested in handy functions and beautiful objects.”

Jean-Paul, a hypermedia author who created the website Cotres.net, wrote in January 2007: “I feel that we are experiencing a ‘floating’ period between the heroic ages, when we were moving forward while waiting for the technology to catch up, and the future, when high-speed broadband will unleash forces that just begin to move, for now only in games.”

The “pervasive” network of the future was described in 2007 by Rafi Haladjian, founder of the internet provider Ozone, on its website: “We will not access the network any more, we will live in it. The future components of this network (wired parts, non-wired parts, operators) will be transparent to the final user. The network will always be open, providing a permanent connection anywhere.”

---

## About encoding

Published in 1963 by the American National Standards Institute (ANSI), ASCII (American Standard Code for Information Interchange) is a 7-bit coded character set for information interchange in English (and Latin). Also named Plain Vanilla ASCII, it is a set of 128 characters with 95 printable unaccented characters (A-Z, a-z, numbers, punctuation and basic symbols).

With computer technology spreading outside North America, the accented characters of a few other languages were included in 8-bit variants of ASCII — also called extended ASCII, ISO-8859 or ISO-Latin — that provided sets of 256 characters, for example ISO 8859-1 (ISO-Latin-1) for French, German and Spanish. But these variants quickly became difficult to handle because of data corruption and conversion issues.

Brian King, director of the Worldwide Language Institute (WWLI), wrote in September 1998 in an email interview: “Computer technology has traditionally been the sole domain of a ‘techie’ elite, fluent in both complex programming languages and in English — the universal language of science and technology. Computers were never designed to handle writing systems that could not be translated into ASCII. There was not much room for anything other than the 26 letters of the English alphabet in a coding system that originally couldn’t even recognise acute accents and umlauts — not to mention non-alphabetic systems like Chinese. But tradition has been turned upside down. Technology has been popularised. GUIs like Windows and Macintosh have hastened the process (and indeed it is no secret that it was Microsoft’s marketing strategy to use their operating system to make computers easy to use for the average person). These days this ease of use has spread beyond the PC to the virtual, networked space of the internet, so that now non-programmers can even insert Java applets into their web pages without understanding a single line of code.”

“An extension of (local) popularisation is the export of information technology around the world. Popularisation has now occurred on a global scale and English is no longer necessarily the lingua franca of the user. Perhaps there is no true lingua franca, but only the individual languages of the users. One thing is certain — it is no longer necessary to understand English to use a computer, nor it is necessary to have a degree in computer science. A pull from non-English-speaking computer users and a push from technology companies competing for global markets has made localisation a fast growing area in software and hardware development. This development has not been as fast as it could have been. The first step was for ASCII to become extended ASCII. This meant that computers could begin to start recognising the accents and symbols used in variants of the English alphabet — mostly used by European languages. But only one language could be displayed on a page at a time.”

“The most recent development [in 1998] is Unicode. Although still evolving and only just being incorporated into the latest software, this new coding system translates each character into 16 bits. Whereas 8-bit extended ASCII could only handle a maximum of 256 characters, Unicode can handle over 65,000 unique characters and therefore potentially accommodate all of the world’s writing systems on the computer. So now the tools are more or less in place. They are still not perfect, but at last we can surf the web in Chinese, Japanese, Korean, and numerous other

languages that don't use the Western alphabet. As the internet spreads to parts of the world where English is rarely used — such as China, for example, it is natural that Chinese, and not English, will be the preferred choice for interacting with it. For the majority of the users in China, their mother tongue will be the only choice.”

First published in January 1991, Unicode provides a unique number for every character, no matter the platform, the program and the language. A double-byte platform-independent encoding allows the processing, storage and interchange of text data in any language. Unicode is maintained by the Unicode Consortium, and is a component of the specifications of the World Wide Web Consortium (W3C), with UTF-8, UTF-16 and UTF-32 variants (UTF: Unicode Transformation Format). Unicode superseded ASCII in December 2007 as the main encoding system on the internet.

---

## About internationalisation

Babel was a joint project created in 1997 by the Internet Society (ISOC) and Alis Technologies, a language software company, to contribute to the internationalisation of the internet. “Towards communicating on internet in any language...” was the home page’s subtitle. Available in seven languages (English, French, German, Italian, Portuguese, Spanish, Swedish), Babel’s website offered a typographical and linguistic glossary for each language. A web page named “The Internet and Multilingualism” explained how to develop a multilingual website, and how to code the world’s writing.

Babel ran the first major study on the distribution of languages on the web. The results were published in June 1997 on a web page named “The Web Languages Hit Parade”, available in seven languages too. English (82.3 percent) was followed by German (4 percent), Japanese (1.6 percent), French (1.5 percent), Spanish (1.1 percent), Swedish (1.1 percent) and Italian (1 percent).

According to Global Reach, the fastest growing language groups creating new websites in 1998 were Spanish (22.4 percent), Japanese (12.3 percent), German (14 percent) and French (10 percent).

Randy Hobler, a marketing consultant for machine translation software and services, wrote in September 1998 in an email interview: “Because the internet has no national boundaries, the organisation of users is bounded by other criteria driven by the medium itself. In terms of multilingualism, you have virtual communities, for example, of what I call ‘language nations’ — all those people on the internet wherever they may be, for whom a given language is their native language. Thus, the Spanish Language nation includes not only Spanish and Latin American users, but millions of Hispanic users in the U.S., as well as odd places like Spanish-speaking Morocco. (...) 85 percent of the content of the web is in English and going down. This trend is driven not only by more websites and users in non-English-speaking countries, but by increasing localisation of company and organisation sites, and increasing use of machine translation to/from various languages to translate websites.”

Yoshi Mikami, a Japanese computer scientist, co-wrote “The Multilingual Web Guide” (with his colleagues Kenji Sekine and Nobutoshi Kohara) on viewing, understanding and creating multilingual web pages. The book was published in Japanese in August 1997 by O’Reilly Japan, before being translated into English, French and German the following year.

Yoshi Mikami explained in December 1998 in an email interview: “My native tongue is Japanese. Because I had my graduate education in the U.S. and worked in the computer business, I became bilingual in Japanese and American English. I was always interested in languages and different cultures, so I learned some Russian, French and Chinese along the way. In late 1995, I created on the web [the page] ‘The Languages of the World by Computers and the Internet’ and tried to summarise there the brief history, linguistic and phonetic features, writing system and computer processing aspects for each of the six major languages of the world, in English and Japanese. As I gained more experience, I invited my two associates to help me write a book on viewing, understanding and creating multilingual web pages, which was published in August 1997 as ‘The Multilingual Web Guide’, in a Japanese edition, the world’s first book on such a subject.”

Yoshi Mikami added: “Thousands of years ago, in Egypt, China and elsewhere, people were more concerned about communicating their laws and thoughts not in just one language, but in several. In our modern world, most nation states have each adopted one language for their own use. I predict greater use of different languages and multilingual pages on the internet (and not a simple gravitation to American English), and also more creative use of multilingual computer translation.”

The World Wide Web Consortium (W3C) offered a web page “Internationalisation / Localisation” in 1998, with the protocols needed to create a bilingual or plurilingual website (HTML, base character set, new tags and attributes, HTTP, language negotiation, URLs, and other identifiers including non-ASCII characters).

---

## About multilingualism

Geoffrey Kingscott, director of the language consultancy Praetorius in London, wrote in September 1998 in an email interview: “Because the salient characteristics of the web are the multiplicity of site generators and the cheapness of message generation, as the web matures it will in fact promote multilingualism. The fact that the web originated in the USA means that it is still predominantly in English but this is only a temporary phenomenon. If I may explain this further, when we relied on the print and audiovisual (film, television, radio, video, cassettes) media, we had to depend on the information or entertainment we wanted to receive being brought to us by agents (publishers, television and radio stations, cassette and video producers) who have to subsist in a commercial world or — as in the case of public service broadcasting — under severe budgetary restraints. That means that the size of the customer base is all important, and determines the degree to which languages other than the ubiquitous English can be accommodated. These constraints disappear with the web.”



Alain Bron, an information systems specialist and writer living in Paris, explained in January 1999: "Different languages will still be used for a long time to come, and this is healthy for the right to be different. The risk is of course an invasion of one language to the detriment of others, and with it the risk of cultural standardisation. I think that online services will gradually emerge to get around this problem. First, translators will be able to translate and revise texts by request. Then, websites with a large audience will provide different language versions, just as the audiovisual industry does now."

Marcel Grangier, head of the French Section of the Swiss Federal Government's Central Linguistic Services, wrote in January 1999: "We can see multilingualism on the internet as a happy and irreversible inevitability. So we have to laugh at the doomsayers who only complain about the supremacy of English. Such supremacy is not wrong in itself, because it is mainly based on statistics (more PCs per inhabitant, more people speaking English, etc.). The answer is not to 'fight' English, much less whine about it, but to build more websites in other languages. As a translation service, we also recommend that websites be multilingual. The increasing number of languages on the internet is inevitable and can only boost multicultural exchanges. For this to happen in the best possible circumstances, we still need to develop tools to improve compatibility. Fully coping with accents and other characters is only one example of what can be done."

According to Bruno Didier, webmaster of the Pasteur Institute Library, interviewed in August 1999: "The internet doesn't belong to any one nation or language. It is a vehicle for culture, and the first vector of culture is language. The more languages there are on the internet, the more cultures will be represented there. I don't think we should give in to the knee-jerk temptation to translate web pages into a largely universal language. Cultural exchanges will only be real if we are prepared to meet with the other culture in a genuine way. And this effort involves understanding the other culture's language. This is very idealistic of course. In practice, when I am monitoring, I curse Norwegian or Brazilian websites where there isn't any English."

Steven Krauwer, coordinator of ELSNET (European Network of Excellence in Human Language Technologies), wrote in September 1998: "As a European citizen I think that multilingualism on the web is absolutely essential, as in the long run I don't think that it is a healthy situation when only those who have a reasonable command of English can fully exploit the benefits of the web; as a researcher (specialised in machine translation) I see multilingualism as a major challenge: how can we ensure that all information on the web is accessible to everybody, irrespective of language differences."

He added in August 1999: "I have become more and more convinced we should be careful not to address the multilingualism issue in isolation. I have just returned from a wonderful summer vacation in France, and even if my knowledge of French is modest (to put it mildly), it is surprising to see that I still manage to communicate successfully by combining my poor French with gestures, facial expressions, visual clues and diagrams. I think the web (as opposed to old-fashioned text-only email) offers excellent opportunities to exploit the fact that transmission of information via different channels (or modalities) can still work, even if the process is only partially successful for each of the channels in isolation."

"Multilingualism could be promoted at the author end, at the server end, and at the browser end. At the author end: better education of web authors to use combinations of modalities to make communication more effective across language barriers (and not just for cosmetic reasons). At the

server end: more translation facilities such as AltaVista (quality not impressive, but always better than nothing). At the browser end: more integrated translation facilities (especially for the smaller languages), and more quick integrated dictionary lookup facilities.”

---

## About website localisation

In late 1997, Yahoo!, then a directory of websites, was the first web portal to offer its home page in seven languages (English, French, German, Japanese, Korean, Norwegian, Swedish) for a growing number of non-English-speaking users.

Brian King, director of the Worldwide Language Institute (WWLI), wrote in September 1998 in an email interview: “Although a multilingual web may be desirable on moral and ethical grounds, such high ideals are not enough to make it other than a reality on a small scale. As well as the appropriate technology being available so that the non-English speaker can go, there is the impact of electronic commerce as a major force that may make multilingualism the most natural path for cyberspace. Sellers of products and services in the virtual global marketplace into which the internet is developing must be prepared to deal with a virtual world that is just as multilingual as the physical world. If they want to be successful, they had better make sure they are speaking the languages of their customers!”

Bill Dunlap has made a life of bringing high-tech products and services to international markets. After graduating from the Massachusetts Institute of Technology (MIT), he created a company to export Apple and PC software to European markets in the early 1980s. He worked as AST Research’s first European sales manager before becoming Compaq’s first sales manager in France. He worked at Compaq’s European headquarters in Munich, Germany, to manage Scandinavian sales. In the mid-1980s, he developed an international marketing consultancy named Euro-Marketing Associates, with two offices in Paris and San Francisco. In 1995, Euro-Marketing Associates was restructured into a virtual consultancy named Global Reach, to help U.S. companies expand their internet presence abroad. This included translating their website into other languages, actively promoting it, and using local online banner advertising to increase local website traffic.

Bill Dunlap wrote in December 1998 in an email interview: “There are so few people in the U.S. interested in communicating in many languages — most Americans are still under the delusion that the rest of the world speaks English. However, in Europe, the countries are small enough so that an international perspective has been necessary for centuries. Since 1981, when my professional life started, I’ve been involved with bringing American companies in Europe. This is very much an issue of language, since the products and their marketing have to be in the languages of Europe in order for them to be visible here. Since the web became popular in 1995 or so, I have turned these activities to their online dimension, and have come to champion European e-commerce among my fellow American compatriots. Most lately at Internet World in New York, I spoke about European e-commerce and how to use a website to address the various markets in Europe.”

He also explained on Global Reach's website: "Promoting your website is at least as important as creating it, if not more important. You should be prepared to spend at least as much time and money in promoting your website as you did in creating it in the first place. After a website's home page is available in several languages, the next step is the development of content in each language. A webmaster will notice which languages draw more visitors (and sales) than others, and these are the places to start in a multilingual web promotion campaign. At the same time, it is always good to increase the number of languages available on a website: just a home page translated into other languages would do for a start, before it becomes obvious that more should be done to develop a certain language branch on a website."

Peter Raggett, head of the library of the Organisation for Economic Cooperation and Development (OECD), wrote in August 1999: "I think it is incumbent on European organisations and businesses to try and offer websites in three or four languages if resources permit. In this age of globalisation and electronic commerce, businesses are finding that they are doing business across many countries. Allowing French, German, Japanese speakers to easily read one's website as well as English speakers will give a business a competitive edge in the domain of electronic trading."

---

## About remote collaboration

Pierre Ruetschi, a journalist for the Swiss daily Tribune de Genève, interviewed Tim Berners-Lee, inventor of the web, for its 20 December 1997 issue. One of his questions was: "Seven years later, are you satisfied with the way the web has evolved?" Tim Berners-Lee answered he was "pleased with the rich and diverse information" available online, but he would like "the web to be more interactive, and people to be able to create information together" because the web was intended as "a medium for collaboration, a world of knowledge that we share".

Christiane Jadelot, a French researcher working at the National Institute of the French Language, wrote in July 1998 about her first steps on the internet, and about the collaborative spirit she found: "I began to really use the internet in 1994, with a browser called Mosaic. I found it a very useful way of improving my knowledge of computers, linguistics, literature... everything. I was finding the best and the worst, but as a discerning user, I had to sort it all out and make choices. I particularly liked the software for email, for file transfers and for dial-up connections. At that time I had problems with a software called Paradox and character sets that I couldn't use. I tried my luck and threw out a question in a specialist news group. I got answers from all over the world. Everyone seemed to want to solve my problem!"

Murray Suid, a writer living in Silicon Valley, wrote in September 1998 in an email interview: "Now, instead of phoning people or interviewing them face to face, I do it via email. Because of speed, it has also enabled me to collaborate with people at a distance, particularly on screenplays. (I've worked with two producers in Germany.) (...) The internet has increased my correspondence dramatically. Like most people, I find that email works better than snail mail. My geographic range of correspondents has also increased — extending mainly to Europe. In the old days, I hardly ever

did transatlantic pen palling. I also find that emailing is so easy, I am able to find more time to assist other writers with their work — a kind of a virtual writing group. This isn't merely altruistic. I gain a lot when I give feedback. But before the internet, doing so was more of an effort."

Robert Ware, a computer scientist, created OneLook Dictionaries in April 1996 as a fast finder in 2 million words from 425 dictionaries (in August 1998) in many fields (business, computer / internet, medical, miscellaneous, religion, science, sports, technology, general, slang). He wrote in September 1998: "I was almost entirely in contact with people who spoke one language and did not have much incentive to expand language abilities. Being in contact with the entire world has a way of changing that. And changing it for the better! I have been slow to start including non-English dictionaries (partly because I am monolingual). But you will now find a few included."

He also wrote about a personal experience: "In 1994, I was working for a college and trying to install a software package on a particular type of computer. I located a person who was working on the same problem and we began exchanging email. Suddenly, it hit me... the software was written only 30 miles away [from Englewood, Colorado] but I was getting help from a person half way around the world. Distance and geography no longer mattered! OK, this is great! But what is it leading to? I am only able to communicate in English but, fortunately, the other person could use English as well as German which was his mother tongue. The internet has removed one barrier (distance) but with that comes the barrier of language."

"It seems that the internet is moving people in two quite different directions at the same time. The internet (initially based on English) is connecting people all around the world. This is further promoting a common language for people to use for communication. But it is also creating contact between people of different languages and creates a greater interest in multilingualism. A common language is great but in no way replaces this need. So the internet promotes both a common language \*and\* multilingualism. The good news is that it helps provide solutions. The increased interest and need is creating incentives for people around the world to create improved language courses and other assistance, and the internet is providing fast and inexpensive opportunities to make them available."

Information on the internet needs to be accessible to all — and can be accessible to all. The French association HandicapZéro launched its first website for visually impaired and blind users in September 2000. HandicapZéro wanted to demonstrate that, "with the respect of some basic rules, the internet can finally become a space of freedom for all". The website quickly became popular with 10,000 visits per month.

The website was revamped in February 2003 with free access to national and international news in real time (in partnership with AFP-Agence France-Presse), sports news (with the daily L'Équipe), TV shows (with the weekly Télérama), weather (with the national service Météo France), as well as a range of services relating to health, employment, leisure and telephony. The website had 2 million visits in 2006.

Blind users can access the website using a Braille device or a speech software. Visually impaired users can set up parameters (size and type of fonts, color of background) on their own "visual profile", that can also be used to read any text document after copying it from the web and pasting

it into the interface. Seeing users can correspond in Braille with blind users through the website, with HandicapZéro providing a free transcription, as well as a free print in Braille sent by postal mail (for free in Europe).

---

## About e-texts

The first e-texts available on the internet were e-zines. As explained by John Labovitz, founder of the E-Zine-List, on its website: “For those of you not acquainted with the zine world, ‘zine’ is short for either ‘fanzine’ or ‘magazine’, depending on your point of view. Zines are generally produced by one person or a small group of people, done often for fun or personal reasons, and tend to be irreverent, bizarre, and/or esoteric. Zines are not ‘mainstream’ publications — they generally do not contain advertisements (except, sometimes, advertisements for other zines), are not targeted towards a mass audience, and are generally not produced to make a profit. An ‘e-zine’ is a zine that is distributed partially or solely on electronic networks like the internet.”

One of the first homes for e-zines was the Etext Archives, founded in 1992 by Paul Southworth, and hosted on the website of the University of Michigan. The Etext Archives were “home to electronic texts of all kinds, from the sacred to the profane, and from the political to the personal”, with six sections: (1) “E-zines”: electronic periodicals from the professional to the personal; (2) “Politics”: political zines, essays, and home pages of political groups; (3) “Fiction”: publications of amateur authors; (4) “Religion”: mainstream and off-beat religious texts; (5) “Poetry”: an eclectic mix of mostly amateur poetry; and (6) “Quartz”: the archive formerly hosted at quartz.rutgers.edu.

As recalled on the website in 1998: “The web was just a glimmer, gopher was the new hot technology, and FTP was still the standard information retrieval protocol for the vast majority of users. The origin of the project has caused numerous people to associate it with the University of Michigan, although in fact there has never been an official relationship and the project is supported entirely by volunteer labor and contributions. The equipment is wholly owned by the project maintainers. The project was started in response to the lack of organised archiving of political documents, periodicals and discussions disseminated via Usenet on newsgroups. Not long thereafter, electronic zines (e-zines) began their rapid proliferation on the internet, and it was clear that these materials suffered from the same lack of coordinated collection and preservation, not to mention the fact that the lines between e-zines (which at the time were mostly related to hacking, phreaking, and internet anarchism) and political materials on the internet were fuzzy enough that most e-zines fit the original mission of The Etext Archives. One thing led to another, and e-zines of all kinds — many on various cultural topics unrelated to politics — invaded the archives in significant volume.”

The E-Zine-List was created by John Labovitz in 1993 as a directory of e-zines around the world (via FTP, gopher, email, the web, and other services), and updated monthly. John Labovitz explained on its website: “I started this list in the summer of 1993. I was trying to find some place to publicise Crash, a print zine I’d recently made electronic versions of. All I could find was the alt.zines newsgroup and the archives at The Well and Etext. I felt there was a need for something less ephemeral and more organised, a directory that kept track of where e-zines could be found. So

I summarised the relevant info from a couple dozen e-zines and created the first version of this list. Initially, I maintained the list by hand in a text editor; eventually, I wrote my own database program (in the Perl language) that automatically generates all the text, links, and files.”

“In the four years I’ve been publishing the list, the net has changed dramatically, in style as well as scale. When I started the list, e-zines were usually a few kilobytes of plain text stored in the depths of an FTP server; high style was having a gopher menu, and the web was just a rumor of a myth. The number of living e-zines numbered in the low dozens, and nearly all of them were produced using the classic self-publishing method: scam resources from work when no one’s looking. Now the e-zine world is different. The number of e-zines has increased a hundredfold, crawling out of the FTP and gopher woodworks to declaring themselves worthy of their own domain name, even to asking for financial support through advertising. Even the term ‘e-zine’ has been co-opted by the commercial world, and has come to mean nearly any type of publication distributed electronically. Yet there is still the original, independent fringe, who continue to publish from their heart, or push the boundaries of what we call a ‘zine’.” The E-Zine-List included 3,045 e-zines on 29 November 1998.

John Mark Ockerbloom, a graduate student at the School of Computer Science (CS) of Carnegie Mellon University (CMU), created The Online Books Page in 1993 as “a website that facilitates access to books that are freely readable over the internet. It also aims to encourage the development of such online books, for the benefit and edification of all.” The web was still in its infancy, with Mosaic as its first browser.

John Mark Ockerbloom wrote in September 1998 in an email interview: “I was the original webmaster here at CMU CS, and started our local web in 1993. The local web included pages pointing to various locally developed resources, and originally The Online Books Page was just one of these pages, containing pointers to some books put online by some of the people in our department. (Robert Stockton had made web versions of some of Project Gutenberg’s texts.) After a while, people started asking about books at other sites, and I noticed that a number of sites (not just Gutenberg, but also Wiretap and some other places) had books online, and that it would be useful to have some listing of all of them, so that you could go to one place to download or view books from all over the net. So that’s how my index got started.”

“I eventually gave up the webmaster job in 1996, but kept The Online Books Page, since by then I’d gotten very interested in the great potential the net had for making literature available to a wide audience. At this point there are so many books going online that I have a hard time keeping up. But I hope to keep up my online books works in some form or another. I am very excited about the potential of the internet as a mass communication medium in the coming years. I’d also like to stay involved, one way or another, in making books available to a wide audience for free via the net, whether I make this explicitly part of my professional career, or whether I just do it as a spare-time volunteer.”

The Online Books Page included 7,000 books in 1998, and began listing serials too. As explained on its website: “Along with books, The Online Books Page is also now listing major archives of serials (such as magazines, published journals, and newspapers), as of June 1998. Serials can be at least as important as books in library research. Serials are often the first places that new research and scholarship appear. They are sources for firsthand accounts of contemporary events and

commentary. They are also often the first (and sometimes the only) place that quality literature appears. (For those who might still quibble about it, back issues of serials are often bound and reissued as hardbound ‘books’.)”

After graduating from Carnegie Mellon University with a PhD in computer science, John Mark Ockerbloom was hired in 1999 as a digital library planner and researcher at the University of Pennsylvania Library. He moved The Online Books Page there, and went on expanding it, with links to 12,000 books in 1999, 20,000 books in 2003 (including 4,000 books published by women), 25,000 books in 2006, 30,000 books in 2008 (including 7,000 books from Project Gutenberg), 35,000 books in 2010, and 2 million books in 2015. The books “have been authored, placed online, and hosted by a wide variety of individuals and groups throughout the world.” The website also provides copyright information in many countries, with links to further reading.

---

## About e-books

Most people who created free online collections of public domain books in the 1990s were inspired by Project Gutenberg, a collection of free e-books created by Michael Hart, a student at the University of Illinois. While Johannes Gutenberg, the inventor of the printing press, had allowed some people to own a few printed books, Project Gutenberg would allow everyone to own a full collection of free e-books.

As recalled by Michael Hart on Project Gutenberg’s website: “When we started [in 1971], the files had to be very small. So doing the U.S. Declaration of Independence (only 5K) seemed the best place to start. This was followed by the Bill of Rights, then the whole U.S. Constitution, as space was getting large (at least by the standards of 1973). Then came the Bible, as individual books of the Bible were not that large, then Shakespeare (a play at a time), and then into general work in the areas of light and heavy literature and references. By the time Project Gutenberg got famous, the standard was 360K disks, so we did books such as ‘Alice in Wonderland’ or ‘Peter Pan’ because they could fit on one disk. Now [in the 1990s] 1.44 is the standard disk and ZIP is the standard compression; the practical file size is about three million characters, more than long enough for the average book.”

Michael Hart wrote in August 1998 in an email interview: “We consider e-text to be a new medium, with no real relationship to paper, other than presenting the same material, but I don’t see how paper can possibly compete once people each find their own comfortable way to e-texts, especially in schools.”

An e-book was a continuous text file instead of a set of pages, using Plain Vanilla ASCII (the low set of ASCII), with caps for the terms in italic, bold or underlined of the print version, for it to be read on any hardware and software. As a text file, a book could be easily copied, indexed, searched, analysed and compared with other books. Books in other languages were also available in extended ASCII to take into account the accentuated characters. Later on, all the books were offered in Unicode, and in several electronic formats.

Project Gutenberg offered e-books in 25 languages in early 2004, e-books in 42 languages (including Sanskrit and the Mayan languages) in July 2005, and e-books in 59 languages in October 2010. At that time, the ten main languages were English (with 28,441 e-books on 7 October 2010), French (1,659 e-books), German (709 e-books), Finnish (536 e-books), Dutch (496 e-books), Portuguese (473 e-books), Chinese (405 e-books), Spanish (295 e-books), Italian (250 e-books) and Greek (101 e-books).

Distributed Proofreaders was created by Charles Franks in 2000 to support the digitisation of public domain books, and quickly became the main source of Project Gutenberg. The books were scanned from a print edition, and converted into text before being proofread by volunteers. Distributed Proofreaders provided a total of 10,000 e-books in December 2006, a total of 20,000 e-books in April 2011, and a total of 30,000 e-books in July 2015, as “unique titles [sent] to the bookshelves of Project Gutenberg, free to enjoy for everybody. Distributed Proofreaders is a truly international community. People from over the world contribute.”

Projekt Runeberg was the first free Swedish collection, and a partner of Project Gutenberg. It was created in 1992 by the computer club Lysator, in cooperation with Linköping University, as a student volunteer project to create and collect free electronic editions of classic Nordic literature. Around 200 e-books were available in 1998. The website also provided a list of 6,000 Nordic authors as a tool for further collection development.

ABU-La Bibliothèque Universelle (ABU-The Universal Library) was the first free French collection, inspired by Project Gutenberg too. It was created in 1993 by ABU (Association of Universal Bibliophiles) to create free electronic editions of classic French literature. 223 e-books from 76 authors were available in 1998.

Projekt Gutenberg-DE was the first free German collection, created in 1994 as a partner of Project Gutenberg. Texts were available for online reading, with one web page for short texts, and several web pages (one per chapter) for longer works. The website also provided an alphabetic list of authors and titles, and a short biography and bibliography for each author.

Athena was the first free Swiss collection, created in 1994 by Pierre Perroud, a teacher at College Voltaire in Geneva, and hosted by the University of Geneva. The bilingual French-English website offered a plurilingual collection in philosophy, science, literature, history and economics. Athena either digitized books (200 titles in December 1997) or provided links to e-books digitised elsewhere (3,500 titles in December 1997, and 8,000 titles one year later). The web page “Helvetia” offered e-books about Switzerland, and the web page “Athena Literature Resources” provided links to other digital collections worldwide.

Pierre Perroud wrote in the February 1996 issue of *Informatique-Informations*, a Swiss computer magazine: “Electronic texts are an encouragement to reading and a convivial participation in the dissemination of culture. They are a good complement to the printed book, which remains irreplaceable for ‘true’ reading. The printed book remains a mysteriously holy companion with a deep symbolism for us: we grip it in our hands, we hold it against us, we look at it with admiration; its small size comforts us and its content impresses us; its fragility holds a density we are fascinated by; like man it fears water and fire, but it has the power to shelter man’s thoughts from time.”



Progetto Manuzio was the first free Italian collection, created in 1995 by Liber Liber, an association “promoting artistic and intellectual creative expression, and using computer technology to link the humanities with the sciences”. Progetto Manuzio was named after the famous 16th-century Venetian publisher who improved the printing techniques invented by Gutenberg. As explained on its website, Progetto Manuzio wanted “to make a noble idea real: the idea of making culture available to everyone. How? By making books, dissertations, articles, fiction — or any other document which could be digitised — available worldwide, at any time, and free of charge. Via a modem or floppy disks (in this case, by adding the cost of a blank disk and postal fees), it is already possible to get hundreds of books. And Progetto Manuzio needs only a few people to make a masterpiece like Dante Alighieri’s ‘La Divina Commedia’ available to millions of people.”

Some collections were organised around an author or several authors, for example The Marx/Engels Internet Archive (MEIA). MEIA was created in 1996 to offer a timeline of the collected works of Karl Marx and Frederick Engels, with links to the digital editions of these works. As explained on its website: “There’s no way to monetarily profit from this project. ‘Tis a labour of love undertaken in the purest communitarian sense. The real ‘profit’ will hopefully manifest in the form of individual enlightenment through easy access to these classic works. Besides, transcribing them is an education in itself. Let us also add that this is not a sectarian / One-Great-Truth effort. Help from any individual or any group is welcome. We have but one slogan: ‘Piping Marx & Engels into cyberspace!’”

The web page “Biographical Archive” offered extensive biographies for Marx and Engels, and short biographies for their family members and friends. The web page “Photo Gallery” offered photos of the Marx and Engels clan from 1839 to 1894, and photos of their dwellings from 1818 to 1895. The web page “Others” listed works by all main Marxist authors, for example James Connolly, Daniel DeLeon and Hal Draper, and a short biography for each. The web page “Non-English Archive” listed works by Marx and Engels that were freely available online in other languages (Danish, French, German, Greek, Italian, Japanese, Polish, Portuguese, Spanish, Swedish). MEIA was later renamed the Marxists Internet Archive.

---

## About the press

The first electronic versions of printed newspapers were available in the early 1990s through commercial services such as America OnLine (AOL) and CompuServe. Newspapers began offering websites in the mid-1990s, with a partial or full version of their latest issue, available for free or with a paid subscription, as well as online archives.

In the United States, the website of The New York Times could be accessed free of charge in 1996, with the content of the printed newspaper, updates with breaking news every ten minutes, and original reporting available only online. The website of The Washington Post provided free daily news and a database of previous articles, with images, sound and video. The website of The Wall Street Journal was available with a paid subscription (100,000 subscribers in 1998).

In the United Kingdom, the daily *The Times* and the weekly *The Sunday Times* were available on the website *Times Online*, with a way to create a personalised edition. The weekly *The Economist* went online too. In other European countries, the daily newspaper *El País* (Spain), the daily newspapers *Le Monde*, *Libération* and *L'Humanité* (France) and the weekly magazines *Der Spiegel* and *Focus* (Germany) were among the first ones to be available online.

Readers could access newspapers that were difficult to find in their own country, for example the Algerian daily *El Watan* on a website created in October 1997. In an article published on 23 March 1998 in the French daily *Le Monde*, Redha Belkhat, chief editor of *El Watan*, explained: "For the Algerian diaspora, to find in London, New York or Ottawa a paper issue of *El Watan* that is less than a week old can be a challenge. Now the newspaper is available in print in Algiers at 6 am, and is available on the internet at noon."

In an essay published in December 1997 on the website of *AJR/NewsLink*, the journalist Eric K. Meyer wrote that "more than 3,600 newspapers now publish on the internet (...). A full 43% of all online newspapers now are based outside the United States. A year ago, only 29% of online newspapers were located abroad. Rapid growth, primarily in Canada, the United Kingdom, Norway, Brazil and Germany, has pushed the total number of non-U. S. online newspapers to 1,563. The number of U.S. newspapers online also has grown markedly, from 745 a year ago to 1,290 six months ago to 2,059 today. Outside the United States, the United Kingdom, with 294 online newspapers, and Canada, with 230, lead the way. In Canada, every province or territory now has at least one online newspaper. Ontario leads the way with 91, Alberta has 44, and British Columbia has 43. Elsewhere in America, Mexico has 51 online newspapers, 23 newspapers are online in Central America and 36 are online in the Caribbean. Europe is the next most wired continent for newspapers, with 728 online newspaper sites. After the United Kingdom, Norway has the next most — 53 — and Germany has 43. Asia (led by India) has 223 online newspapers. South America (led by Bolivia) has 161 and Africa (led by South Africa) has 53. Australia and other islands have 64 online newspapers."

Kushal Dave, a high school freshman at Yale University, wrote in 1998 in an essay about the future of publishing: "It is now possible to read Associated Press Reports as they are released, not in the next morning's paper, and you don't even have to pay the 25 cents. Cost, speed, and availability are just some of the compelling arguments for electronic publishing instead of paper. There are also many companies attempting to capitalise on the multimedia possibilities of electronic publishing. Sound and pictures are being incorporated in low-cost Internet World Wide Web 'publications', and companies like Medio and Nautilus are producing CD-ROMs that represent the new generation of periodicals — now music reviews include sound clips, movie reviews include trailers, book reviews include excerpts, and how-to articles include demonstrative videos. All this is put together with low costs, high speed, and many advantages."

"The internet is certainly a more accessible and convenient medium, and thus it would be better in the long run if the strengths of the print media could be brought online without the extensive costs and copyright concerns that are concomitant. As the transition is made, the neat thing is a growing accountability for previously relatively irreproachable edifices. For example, we already see email addresses after articles in publications, allowing readers to pester authors directly. Discussion forums on virtually all major electronic publications show that future is providing not just one person's opinion but interaction with those of others as well. Their primary job is the provision of background information. Also, the detailed statistics can be gleaned about interest in an

advertisement or in content itself will force greater adaptability and a questioning of previous beliefs gained from focus groups. This means more finely honed content for the individual, as quantity and customisability grows.”

Online newsletters skipped the cost of a print version. They were sent by email or available on a website. As recalled in July 1999 by Jacques Gauchey, a French journalist and writer, in an email interview: “In 1996 I published a few issues of a free English newsletter on the internet. It had about ten readers per issue until the day when the electronic version of Wired created a link to it. In one week I got about 100 emails, some from French readers of my book ‘La Vallée du Risque: Silicon Valley’ [The Valley of Risk: Silicon Valley, published by Plon in 1990], who were happy to find me again.”

---

## About a shift in jobs

In the press industry, articles went directly from text to layout, without being typed anymore by the production staff because journalists were typing their articles themselves on their computer. In the book industry, digitisation sped up the editorial process, which used to be sequential, by allowing the copy editor, the image editor and the layout staff to work on the same book at the same time.

The International Labour Office (ILO) held its first Symposium on Multimedia Convergence in January 1997 in Geneva, Switzerland, “to stimulate reflection on the policies and approaches most apt to prepare our societies and especially our work forces for the turbulent transition towards an information economy.” Discussions between employers, unionists and government representatives from all over the world focused on three main points: (1) the information society: what it meant for governments, employers and workers; (2) the convergence process: its impact on employment and work; (3) labour relations in the information age.

Wilfred Kiboro, managing director of The Nation Printers and Publishers in Kenya, explained that, in his country, each newspaper copy was read by at least twenty people. Distribution costs could drop with the use of a printing system by satellite, avoiding the need to deliver newspapers by truck throughout the country. “Information technology needs to be brought to affordable levels. I have a dream that perhaps in our lifetime in Africa, we will see villagers being able to access the internet from their rural villages where today there is no water and no electricity.”

There were also cultural issues. Many African newspapers, radio programs and TV channels belonged to a few western media groups that gave their own vision of Africa, without fully understanding its economic and social issues. And there were uncertainties about the work itself. “In content creation in the multimedia environment, it is very difficult to know who the journalist is, who the editor is, and who the technologist is that will bring it all together. At what point will telecom workers become involved as well as the people in television and other entities that come to create new products? Traditionally in the print media, for instance, we had printers, journalists, sales and marketing staff and so on, but now all of them are working on one floor from one desk.”

Bernie Lunzer, secretary-treasurer of the Newspaper Guild in the United States, shared its own issues: "Our reporters have seen new deadline pressures build as the material is used throughout the day, not just at the end of the day. There is also a huge safety problem in the newsrooms themselves due to repetitive strain injuries. Some people are losing their careers at the age of 35 or 40, a problem that was unheard of in the age of the typewriter. But as people work 8-to 10-hour shifts without ever leaving their terminals, this has become an increasing problem."

Carlos Alberto de Almeida, president of the National Federation of Journalists (FENAJ) in Brazil, shared similar concerns: "Technology offers the opportunity to rationalise work, to reduce working time and to encourage intellectual pursuits and even entertainment. But so far none of this has happened. On the contrary, media professionals — whether executives, journalists or others — are working longer and longer hours. If one were to rigorously observe the labour legislation and the rights of professionals, then the extraordinary positive aspects of these new technologies would emerge. This has not been the case in Brazil. Journalists can be easily phoned on weekends to do extra work without extra pay."

Some participants, mostly employers, explained that the information society was generating or would generate jobs, while other participants, mostly unionists, explained that unemployment was rising worldwide.

Etienne Reichel, acting director of VisCom (Visual Communication) in Switzerland, explained: "The work of 20 typesetters is now carried out by six qualified workers. There has also been a concentration of centres of production, thus placing enormous pressure on the small and medium-sized enterprises which are traditional sources of employment. Computer science makes it possible for experts to become independent producers. Approximately 30 per cent of employees have set up independently and have been able to carve out part of the market."

Peter Leisink, associate professor of labour studies at Utrecht University, Netherlands, explained: "A survey of the United Kingdom book publishing industry showed that proofreaders and editors have been externalised and now work as home-based teleworkers. The vast majority of them have entered self-employment, not as a first-choice option, but as a result of industry mergers, relocations and redundancies. These people should actually be regarded as casualised workers, rather than as self-employed, since they have little autonomy and tend to depend on only one publishing house for their work."

Michel Muller, secretary-general of the French Federation of the Book, Paper and Communication Industries (FILPAC), stated that jobs in these industries fell from 110,000 to 90,000 in ten years (1987-96), with expensive social plans to retrain and reemploy the 20,000 people who lost their jobs. "If the technological developments really created new jobs, as had been suggested, then it might have been better to invest the money in reliable studies about what jobs were being created and which ones were being lost, rather than in social plans which often created artificial jobs. These studies should highlight the new skills and qualifications in demand as the technological convergence process broke down the barriers between the printing industry, journalism and other vehicles of information. Another problem caused by convergence was the trend towards ownership concentration. A few big groups controlled not only the bulk of the print media, but a wide range of other media, and thus posed a threat to pluralism in expression. Various tax advantages enjoyed by the press today should be re-examined and adapted to the new realities

facing the press and multimedia enterprises. Managing all the social and societal issues raised by new technologies required widespread agreement and consensus. Collective agreements were vital, since neither individual negotiations nor the market alone could sufficiently settle these matters.”

Walter Durling, director of AT&T Global Information Solutions in the United States, had a quite theoretical statement on the matter: “Technology would not change the core of human relations. More sophisticated means of communicating, new mechanisms for negotiating, and new types of conflicts would all arise, but the relationships between workers and employers themselves would continue to be the same. When film was invented, people had been afraid that it could bring theatre to an end. That has not happened. When television was developed, people had feared that it would do away cinemas, but it had not. One should not be afraid of the future. Fear of the future should not lead us to stifle creativity with regulations. Creativity was needed to generate new employment. The spirit of enterprise had to be reinforced with the new technology in order to create jobs for those who had been displaced. Problems should not be anticipated, but tackled when they arose.”

In fact, people were less afraid of technology than they were afraid of losing their jobs. In 1997, unemployment was already significant anywhere, which was not the case when film and television were invented. Unions were struggling worldwide to promote the creation of jobs through investment, innovation, vocational training, computer literacy, retraining in digital technology, fair conditions for labour contracts and collective agreements, fair conditions for the reuse of articles from the print media to the web, protection of workers in the art field, and defence of teleworkers as workers having full rights.

Despite the unions’ efforts, would the situation become as tragic as what was suggested in a note of the symposium’s proceedings? “Some fear a future in which individuals will be forced to struggle for survival in an electronic jungle. And the survival mechanisms which have been developed in recent decades, such as relatively stable employment relations, collective agreements, employee representation, employer-provided job training, and jointly funded social security schemes, may be sorely tested in a world where work crosses borders at the speed of light.”

---

## About copyright

A major issue was the handling of copyright on the internet. Bernie Lunzer, secretary-treasurer of the Newspaper Guild in the United States, explained during the same symposium: “There is a huge battle over intellectual property rights, especially with freelancers, but also with our members who work under collective bargaining agreements. The freelance agreements that writers are asked to sign are shocking. Bear in mind that freelance writers are paid very little. They turn over all their future rights — reuse rights — to the publisher and very little in exchange. Publishers are fighting for control and ownership of product, because they are seeing the future.”

Heinz-Uwe Rübenach, president of the Newspapers Publishers Association in Germany, explained: “Copyright is one of the keys to the future information society. If a publishing house which offers the journalist work, even on an online service, is not able to manage and control the use of the resulting product, then it will not be possible to finance further investments in the

necessary technology. Without that financing, the future becomes less positive and jobs can suffer. If, however, publishers see that they are able to make multiple use of their investment, then obviously this is beneficial for all. Otherwise the costs associated with online services would increase considerably. As far as the European market is concerned, this would only increase competitive pressures, since publishers in the United States do not have to pay for multiple uses.”

In “Intellectual Property Rights and the World Wide Web”, an article published in 1997 on the website of AJR/NewsLink, Penny Pagano, a freelance author, wrote: “Today, those who create information and those who publish, distribute and repackage it are finding themselves at odds with each other over the control of electronic rights. (...) ‘The electronic explosion has changed the entire nature of the business,’ Carlinsky [vice president of the American Society of Journalists and Authors] says. In the past, articles sold to a periodical essentially ‘turned into a pumpkin with no value’ once they were published. ‘But the electronic revolution has extended the shelf life of content of periodicals. You can now take individual articles and put them into a virtual bookstore or put them on a virtual newsstand.’ The second major change in recent years, he says, is ‘an increasing trend to more and more publications being owned by fewer larger and larger companies that tend to be international media conglomerates. They are connected corporately with an enormous array of enterprises that might be interested in secondary use of materials.’ The National Writers Union has created a new agency called the Publication Rights Clearinghouse (PRC). Based on the music industry’s ASCAP [American Society of Composers, Authors, and Publishers], PRC will track individual transactions and pay out royalties to writers for secondary rights for previously used articles. For \$20, freelance writers who have secondary rights to previously published articles can enrol in PRC. These articles become part of a PRC file that is licensed to database companies.”

A major blow for collections of e-books based in the United States was the Digital Millennium Copyright Act (DMCA), signed in October 1998 in the wake of the WIPO Copyright Treaty (WCT) signed in December 1996 by the member states of the World Intellectual Property Organisation (WIPO) to “deal with the protection of works and the rights of their authors in the digital environment”. A number of books supposed to enter the public domain stayed copyrighted.

As explained in October 1998 by John Mark Ockerbloom, founder of The Online Books Page, on its website: “The copyright extension bill will prevent books published in 1923 and later that are not already in the public domain from entering the public domain in the United States for at least 20 years. I have started a page to provide access to copyright renewal records, which eventually should make it easier to find books published after 1922 that have entered the public domain due to non-renewal.”

John Mark Ockerbloom also explained in August 1999 in an email interview: “I think it is important for people on the web to understand that copyright is a social contract that is designed for the public good — where the public includes both authors and readers. This means that authors should have the right to exclusive use of their creative works for limited times, as is expressed in current copyright law. But it also means that their readers have the right to copy and reuse the work at will once copyright expires. In the U.S. now, there are various efforts to take rights away from readers, by restricting fair use, lengthening copyright terms (even with some proposals to make them perpetual) and extending intellectual property to cover facts separate from creative works (such as found in the ‘database copyright’ proposals). There are even proposals to effectively replace copyright law altogether with potentially much more onerous contract law. Stakeholders in

this debate have to face reality, and recognise that both producers and consumers of works have legitimate interests in their use. If intellectual property is then negotiated by a balance of principles, rather than as the power play it too often ends up being ('big money vs. rogue pirates'), we may be able to come up with some reasonable accommodations."

Project Gutenberg created the page "Copyright HowTo" for the volunteers digitising books for its collection. It could be summarised as such: (1) works published before 1923 entered the public domain no later than 75 years from the copyright date: all these works belong to public domain; (2) works published between 1923 and 1977 retain copyright for 95 years: no such works will enter the public domain until 2019; (3) works created from 1978 on enter the public domain 70 years after the death of the author if the author is a natural person: nothing will enter the public domain until 2049; (4) works created from 1978 on enter the public domain 95 years after publication or 120 years after creation if the author is a corporate one: nothing will enter the public domain until 2074.

Copyright legislation became more restrictive too in Europe. The European Union Copyright Directive (Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society) was ratified by the European Commission in May 2001. A copyright based on "author's life plus 70 years" replaced the copyright based on "author's life plus 50 years" in use in most European countries, following pressure from major content owners who successfully lobbied for the harmonisation of national copyright laws as a response to the globalisation of the market. All member countries were required to adapt their own copyright legislation accordingly within a given time frame. In its article 43, the same directive asked all member states to adopt measures for handicapped users to be able to access books, and to promote accessible formats.

---

## About creative copyright

The authors I was interviewing in the late 1990s shared their thoughts on the matter.

According to Jacques Gauchey, a journalist and writer living in San Francisco: "Copyright in its traditional context doesn't exist anymore. Authors have to get used to a new situation: the total freedom of the flow of information. The original content is like a fingerprint: it can't be copied. So it will survive and flourish." (July 1999)

According to Alain Bron, an information systems consultant and writer living in Paris: "I consider the web today as a public domain. That means in practice that the notion of copyright on it disappears: everyone can copy everyone else. Anything original risks being copied at once if copyrights are not formally registered or if works are available without payment. A solution is to make people pay for information, but this is no watertight guarantee against it being copied." (November 1999)

According to Guy Antoine, founder of Windows on Haiti, a website about the Haitian language and culture: “The debate will continue forever, as information becomes more conspicuous than the air that we breathe and more fluid than water. Authors will have to become a lot more creative in terms of how to control the dissemination of their work and profit from it. The best that we can do right now is to promote basic standards of professionalism, and insist at the very least that the source and authorship of any work be duly acknowledged. Technology will have to evolve to support the authorisation process.” (November 1999)

Some people worked on creative solutions to adapt copyright to the web, such as the copyleft, the GPL (GNU General Public License) and the Creative Commons licenses.

The term “copyleft” was invented in 1984 by Richard Stallman, a software developer at the Massachusetts Institute of Technology (MIT), who created the Free Software Foundation (FSF) and the GNU Project, a collaborative project for the development of free software. As explained on its website: “Copyleft is a general method for making a program or other work free, and requiring all modified and extended versions of the program to be free as well. Copyleft says that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it. Copyleft guarantees that every user has freedom. Copyleft is a way of using the copyright on the program. It doesn’t mean abandoning the copyright; in fact, doing so would make copyleft impossible. The word ‘left’ in ‘copyleft’ is not a reference to the verb ‘to leave’ — only to the direction which is the inverse of ‘right’.”

“The GNU General Public License (GPL) is a free, copyleft license for software and other kinds of works. The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program — to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too. (...) The GNU Free Documentation License (GFDL) is a form of copyleft intended for use on a manual, textbook or other document to assure everyone the effective freedom to copy and redistribute it, with or without modifications, either commercially or non-commercially.”

Creative Commons (CC) was founded in 2001 by Lawrence Lessig, a professor of law at the Stanford Law School in California, to offer free licenses for authors to be able to share their work on the internet, and for creators to be able to copy and remix it. As explained on its website: “Creative Commons is a nonprofit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. We provide free licenses and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof.”

Who has used the Creative Commons licenses? O’Reilly Media for its books, Wikipedia for its articles, and the Public Library of Science (PLOS) for its journals, for example. There were one million Creative Commons licensed works in 2003, 4.7 million licensed works in 2004, 20 million licensed works in 2005, 50 million licensed works in 2006, 90 million licensed works in 2007, 130 million licensed works in 2008, 400 million licensed works in 2010, and 882 million licensed works in 2014.

---



## About bookstores

The Internet Bookshop (IBS), which was the largest online bookstore in Europe with 1.4 million titles in 1997, experimented innovative solutions that would later inspire Amazon and others.

The Internet Bookshop developed a system of affiliates in January 1997. Any website owner could recommend books and sell them, with the Internet Bookshop handling secure online ordering, customer service, shipping, and weekly sales reports by email. The affiliates earned 10 percent of the sales — with the need for a new legislation for earnings made through the internet. Affiliates ranged from readers or authors to publishers, businesses and nonprofits. Amazon created its own Associates Program in June 1997.

The Internet Bookshop began selling books published in the United States to its own clients in September 1997, followed by Waterstones in 1998. The Publishers Association in the U.K. tried to stop them, and also banned online bookstores based in the U.S. from selling books to customers based in the U.K.

The Internet Bookshop began offering significant discounts in October 1997, with discounts up to 45 percent for some best-sellers, followed by lawsuits from physical bookstores and publishers. On its website, the page IBS News gave daily news on the frictions, debates, complaints and tense negotiations between online bookstores and physical bookstores, between online bookstores and associations of publishers, and between the online bookstores themselves, that were claiming priority for the customers living in the country where they were based.

The Internet Bookshop also led the fight to remove national borders for selling books, including for taxation. Customers began buying books across borders, and the legislation followed. An outline agreement was concluded in December 1997 between the European Union and the United States, and followed by an international convention. The internet became a free trade area for books, movies and software bought online. Online goods (books, CDs, DVDs) and services were subject to existing regulations, with the collection of VAT, but with no additional custom taxes.

Amazon's two first subsidiaries outside the United States were created in October 1998 in the United Kingdom and in Germany. Amazon had 1.8 million clients in the United Kingdom, 1.2 million clients in Germany, and less than 1 million clients in France in August 2000, when it opened Amazon France, with books, music, DVDs and videos (and software and video games from June 2001) and a 48-hour delivery.

Unlike their counterparts in the United States and in the United Kingdom, where book prices were free, French online bookstores couldn't offer significant bargains. Prices were regulated by the Lang Law, fathered by French minister of culture Jack Lang to protect independent bookstores. The 5 percent discount allowed to any physical or online bookstore offered little latitude to online bookstores, that were nevertheless optimistic about the prospects of e-commerce, with online sales only 0.5 percent of the book market (5.4 percent in the U.S.).

Amazon's economic model was admired by many in Europe, but could hardly be considered a model for staff management, with short-term labour contracts, low wages and poor working conditions. Despite the secrecy surrounding the working conditions of Amazon's employees, problems began to filter. In November 2000, after meeting with 50 employees from Amazon's French distribution center in Boigny-sur-Bionne, the U.K.-based Prewitt Organizing Fund and the French union SUD-PTT Loire-Atlantique started an awareness campaign among Amazon France's employees. In a statement following the meeting, SUD-PTT Loire Atlantique reported "degraded working conditions, flexible schedules, short-term labour contracts in periods of flux, low wages, and minimal social guarantees". Similar actions took place in Germany and in the United Kingdom. Patrick Moran, head of the Prewitt Organising Fund, founded an employee organisation named Alliance of New Economy Workers. In response, Amazon sent internal memos to its employees, stating the pointlessness of unions within the company.

After a deficit in the fourth quarter 2000, Amazon reduced its workforce by 15 percent in January 2001, and 270 employees lost their jobs in Europe (1,300 employees lost their jobs in the U.S.). Amazon closed its customer service center in The Hague, Netherlands, and its 240 employees were offered to relocate in another customer service center in Slough, United Kingdom, or in Regensburg, Germany. Amazon diversified its products sold online, with only 58 percent of the sales for cultural products in November 2001. For its 10th anniversary in July 2005, Amazon had 9,000 employees and 41 million customers worldwide for a whole range of products with a 48 hour-delivery in one of the seven countries (United States, United Kingdom, Germany, France, Japan, China, Canada) with an Amazon distribution center.

What about small bookshops? Local brick-and-mortar bookshops closed one after the other, but some of them fought back, like Librairie Ulysse (Ulysses Bookshop) in Paris, a travel bookshop created in 1971 by Catherine Domain. Its 20,000 out-of-print or new books, periodicals and maps about any country are packed up in a tiny space in the heart of Paris, on the famed Île Saint-Louis surrounded by the river Seine.

Catherine Domain has recounted her first steps as a bookseller on the website of Librairie Ulysse: "After traveling for ten years on every continent, I stopped and told myself: 'What am I going to do for a living?' I was aware I needed to be part of society in one way or another. I made a choice by deduction, because I didn't want to have an employer or an employee. I remembered my grandfathers; one was a navigator, and the other one was a bookseller in Périgord [in southern France]. I also realised that I needed to visit more than a dozen bookshops before finding any documentation on a country as close as Greece. So a travel bookshop came to my mind during a world tour, while I was sailing between Colombo and Surabaya."

"Back in Paris, I looked for a shop, inquired about being a bookseller, worked as an intern in other bookshops, wrote index cards, and thought about a name for my new business. One morning, when I was on my way to buy my daily newspaper, I looked up and saw the shop 'Ulysse' — a reference to Joyce — at number 35 in the street Saint-Louis-en-Île. 'Here is a good name!', I told myself. I climbed two stairs and entered a very small 16m<sup>2</sup> [172 square feet] shop. Four guys were playing poker. 'What a cute shop!', I said. 'It is for sale', one player said without even looking up. 48 hours later, I was a bookseller. This was in September 1971. The first travel bookshop in the world was born."

“Twenty years later, my shop didn’t survive the real estate development in the heart of Paris and its rocket prices, and I had to move out, like many people in my neighbourhood. Luckily, my stubborn side — I am a Taurus ascendant Taurus — gave me the strength to move my bookshop a few meters away in a larger place, at number 26 in the street Saint-Louis-en-l’Île, in a quite uncommon building, for two reasons. First, this was the building where I first lived in Île Saint-Louis a long time ago. Second, this building formerly hosted the branch of a bank that was famously burglarised by Spaggiari.”

Catherine Domain sailed around the globe and visited 140 countries, with a few challenging trips. But her most difficult challenge was to create the bookshop’s website on her own, from scratch, without knowing anything about computers. She wrote in December 1999 in an email interview: “My first year with a computer and the internet was one long technical agony! My site is still pretty basic and under construction. Like my bookshop, it is a place to meet people before being a place to do business. The internet is a pain in the neck, takes a lot of my time and I earn hardly any money from it, but that doesn’t worry me. I am very pessimistic though, because the internet is killing off specialist bookshops.”

However she created a second travel bookshop in 2005, this time facing the ocean, in Hendaye, near the Spanish frontier. At high tide, the bookshop is like a boat of books, and its floor is sometimes flooded by the sea. It is open from 20 June to 20 September, while her partner, an old maps specialist, runs her bookshop in Paris.

Catherine Domain wrote in April 2010: “The internet has taken more and more space in my life! I have become a publisher after some painful training to learn how to use Photoshop and InDesign. This is also a great joy to see that the political will to keep people in front of their computers — for them not to start a revolution — can be defeated by giant spontaneous happy hours [organised through Facebook] with thousands of people who want to speak with one another in person. In the end, there will always be unexpected developments to new inventions. When I started using the internet, I didn’t expect at all to become a publisher.”

---

## About e-bookstores

Online bookstores began selling e-books. These e-bookstores were either digital bookstores specializing in e-books (for example Palm Digital Media, Yahoo! eBookStore, Adobe Digital Media Store, Mobipocket eBookStore and Numilog) or part of an online bookstore that also sold printed books (for example Amazon and Barnes & Noble.com).

Denis Zwirn, founder and head of Numilog, wrote in August 2007 in an email interview: “E-books are not a topic for symposiums, conceptual definitions or divination by some ‘experts’ any more. They are a commercial product and a tool for reading. We need to offer books that can be easily read on any electronic device used by our customers, sooner or later with an electronic ink display. And to offer them as an industry. The digital book is not — and will never be — a niche product (dictionaries, travel guides, books for blind users). It is becoming a mass market product, with multiple forms, like the traditional book.”

Mobipocket was founded in March 2000 by Thierry Brethes and Nathalie Ting as a company specialising in e-books for PDAs. The Mobipocket format (PRC) and the Mobipocket Reader could be used on any PDA in 2000 and on any computer in 2002. The Mobipocket Reader was available in five languages (French, English, German, Spanish, Italian) in 2003. 6,000 titles in several languages were sold on Mobipocket's eBookStore or on partners' online bookstores. Mobipocket (format, software and e-books) was bought by Amazon in 2005 before it launched the Kindle in November 2007.

The first e-books were available as PDF files. A veteran format created in June 1993, PDF was perfected over the years as a global standard for information distribution and viewing. Acrobat Reader and Adobe Acrobat gave the tools to view and create PDF files in several languages and for several platforms (Windows, Mac, Linux, Palm OS, Pocket PC, Symbian OS, and others). Adobe opened its Digital Media Store in late 2003, with titles in PDF from major publishers (HarperCollins, Random House, Simon & Schuster) as well as newspapers and magazines (The New York Times, Popular Science, etc.). Adobe eBooks Central was created as a service to read, publish, sell and lend e-books. The Adobe eBook Library was created as a prototype digital library. After being a proprietary format, PDF was officially released as an open standard in July 2008, and published by the International Organisation for Standardisation (ISO) as ISO 32000-1:2008.

With so many proprietary formats showing up in the late 1990s, the digital publishing industry worked on a standard format, and released in September 1999 the Open eBook (OeB), based on XML and defined by OeBPS (Open eBook Publication Structure). The Open eBook Forum was created in January 2000 to develop the OeB format and the OeBPS specifications. Since then, most e-book formats have derived from — or are compatible with — the OeB format, for example the PRC format (Mobipocket) or the LIT format (Microsoft). The Open eBook Forum became the International Digital Publishing Forum (IDPF) in April 2005. The OeB format was replaced in 2007 by the EPUB (Electronic Publication) format as a global standard for e-books, designed for reflowable content on any device.

PDF or EPUB? Marc Autret, a developer and graphic designer, wrote in April 2011 in an email interview: "I do regret that the emergence of EPUB has led to the obliteration of PDF as a format for e-books. The fact that interactivity within a PDF cannot be displayed by the current mobile platforms has removed any possibility of experimenting new things that had seemed very promising to me. While print publishing offers many different objects, ranging from the carefully designed art book to the basic book for everyday reading, the e-book market has grown from the start on a totalitarian and segregationist mode, comparable to a war between operating systems, instead of favouring a technical and cultural emulation. We now see few PDF e-books exploring the opportunities given by this format."

"In the unconscious collective mind, PDF has stayed a kind of static duplicate of the printed book, and nobody wants to see anything else. The EPUB format, which is nothing but a combination of XHTML/CSS (admittedly with JavaScript prospects), consists in putting e-books 'in phase with' the web. This technology has favoured structured content, but hasn't favoured typographic craft at all. It has given a narrow vision of digital work, reducing it to a flow of information. We don't measure it yet, but the worst cultural disaster in recent decades has been the advent of XML as a language that pre-calibrates and contaminates the way we think our hierarchies. XML and its avatars continue to lock us in the cultural invariants of the western world."

---

## About authors

Murray Suid, a writer living in Silicon Valley, was among the first authors to offer a web extension to his educational books, children books and screenplays. He wrote in September 1998 in an email interview: “In a time of great change, many ‘facts’ don’t stay factual for long. In other words, many books go quickly out of date. But if a book can be web-extended (living partly in cyberspace), then an author can easily update and correct it, whereas otherwise the author would have to wait a long time for the next edition, if indeed a next edition ever came out. I do not know if I will publish books on the web — as opposed to publishing paper books. Probably that will happen when books become multimedia. (I am currently helping develop multimedia learning materials, and it is a form of teaching that I like a lot — blending text, movies, audio, graphics, and when possible interactivity).”

“Also, in terms of marketing, the web seems crucial, especially for small publishers that can’t afford to place ads in major magazines and on the radio. Although large companies continue to have an advantage, in cyberspace small publishers can put up very competitive marketing efforts. We think that paper books will be around for a while, because using them is habitual. Many readers like the feel of paper, and the ‘heft’ of a book held in the hands or carried in a purse or backpack. I haven’t yet used a digital book, and I think I might prefer one — because of ease of search, because of colour, because of sound, etc. Obviously, multimedia ‘books’ can be easily downloaded from the web, and such books probably will dominate publishing in the future. Not yet though. I would also like to have direct access to text — digitally read books in the Library of Congress, for example, just as now I can read back issues of many newspapers. Currently, while I can find out about books online, I need to get the books into my hands to use them. I would rather access them online and copy sections that I need for my work, whereas today I either have to photocopy relevant pages, or scan them in, etc.”

Murray Suid added in August 1999: “In addition to ‘web-extending’ books, we are now web-extending our multimedia (CD-ROM) products — to update and enrich them.” He added in October 2000: “Our company — EDVantage Software — has become an internet company instead of a multimedia (CD-ROM) company. We deliver educational material online to students and teachers.”

The internet is a character in itself in Alain Bron’s second novel “Sanguine sur Toile” (Sanguine on the Net), available in print from Éditions du Choucas in 1999, and as an e-book (PDF) from Éditions 00h00 in 2000.

Alain Bron wrote in November 1999 in an email interview: “In my novel, the internet is a character in itself. Instead of being described in its technical complexity, it is depicted as a character that can be either threatening, kind or amusing. Remember the computer screen has a dual role — displaying as well as concealing. This ambivalence is the theme throughout. In such a game, the big winner is of course the one who knows how to free himself from the machine’s grip and put humanism and intelligence before everything else.”

“My novel is the strange story of an internet user caught up in an upheaval inside his own computer, which is being remotely operated by a mysterious person whose only aim is revenge. I wanted to take the reader into the worlds of painting and business, which intermingle, escaping and meeting up again in the dazzle of software. The reader is invited to try to untangle for himself the threads twisted by passion alone. To penetrate the mystery, he will have to answer many questions. Even with the world at his fingertips, isn't the internet user the loneliest person in the world? As for competition, what is the greatest degree of violence possible in a company these days? Does painting tend to reflect the world or does it create another one? I also wanted to show that images are not that peaceful. You can use them to take action, even to kill.”

Alain Bron wrote in the same email interview: “I spent about 20 years at Bull. There I was involved in all the adventures of computer and telecommunications development. I was a representative of the computer industry at ISO and chaired the network group of the X/Open consortium. I also took part in the early internet with my colleagues of Honeywell in the U.S. in late 1978. I am now an information systems consultant, and I keep the main computer systems of major companies and their foreign subsidiaries running smoothly. And I write. I have been writing since I was a teenager. Short stories (about 100), psycho-sociological essays, articles and novels. It is an inner need as well as a great pleasure.”

“The important thing in the internet is the human value that is added to it. The internet can never be shrewd about a situation, take a risk or replace the intelligence of the heart. The internet simply speeds up the decision-making process and reduces uncertainty by providing information. We still have to leave time to time, let ideas mature and bring a touch of humanity to a relationship. For me, the aim of the internet is meeting people, not increasing the number of electronic exchanges.”

In Spain, Arturo Pérez-Reverte was the best-selling author of *Alatriste*, a series of novels about the adventures of Capitan Alatriste in the 17th century. After three titles published in print in 1997, 1998 and 1999, the new title to be released in late 2000 was “*El Oro del Rey*” (“The King's Gold”). Arturo Pérez-Reverte partnered with his publisher Alfaguara in November 2000 to publish “*El Oro del Rey*” as a PDF that could be downloaded from the web portal *Inicia* for one month, before the release of the printed book in physical bookstores in December 2000. The PDF could be downloaded for 2.90 euros, a cheap price compared to the 15.10 euros for the printed book. One month later, there were 332,000 downloads, but only 12,000 readers who paid for them. Most readers shared their password on chat forums. While this digital experiment was not a financial success, it was a great marketing tool to launch the printed book.

Jean-Paul, a writer living in Paris, switched from being a print author to being a hypermedia author, and searched how hyperlinks could expand his writing towards new directions on his website *Cotres.net*.

Jean-Paul wrote in June 2000 in an email interview: “The internet allows me to do without intermediaries, such as record companies, publishers and distributors. Most of all, it allows me to crystallise what I have in my head: the print medium (desktop publishing, in fact) only allows me to partly do that. Surfing the web is like radiating in all directions (I am interested in something and I click on all the links on a home page) or like jumping around (from one click to another, as the links appear). You can do this in the written media, of course. But the difference is striking. So the internet didn't change my life, but it did change how I write. You don't write the same way for a website as you do for a script or a play.”

“In fact, it wasn’t exactly the internet that changed my writing, it was the first model of the Mac. I discovered it when I was teaching myself HyperCard. I still remember how astonished I was during my month of learning about buttons and links and about surfing by association, objects and images. Being able, by just clicking on part of the screen, to open piles of cards, with each card offering new buttons and each button opening onto a new series of them. In short, learning everything about the web that today seems really routine was a revelation for me. I have heard that Steve Jobs and his team had a similar shock when they discovered the forerunner of the Mac in the laboratories of Rank Xerox.”

“Since then I have been writing directly on the screen. I use a paper print-out only occasionally, to help me fix up an article, or to give somebody who doesn’t like screens a rough idea, something immediate. It is only an approximation, because print forces us into a linear relationship: the words scroll out page by page most of the time. But when you have links, you have got a different relationship to time and space in your imagination. And for me, it is a great opportunity to use this reading/writing interplay, whereas leafing through a book gives only a suggestion of it — a vague one because a book is not meant for that.”

Jean-Paul insisted on the growing interaction between cyber-literature and technology: “The future of cyber-literature, techno-literature or whatever you want to call it, is set by the technology itself. It is now impossible for an author to handle all by himself the words and their movement and sound. A decade ago, you could know well each of Director, Photoshop or Cubase (to cite just the better-known software), using the first version of each. That is not possible any more. Now we have to know how to delegate, find more solid financial partners than Gallimard, and look in the direction of Hachette-Matra, Warner, the Pentagon and Hollywood. At best, the status of the, what... multimedia director? will be the status of video director, film director, the manager of the product. He is the one who receives the golden palms at Cannes, but who would never have been able to earn them just on his own. As twin sister (not a clone) of the cinematograph, cyber-literature (video + the link) will be an industry, with a few isolated craftsmen on the outer edge (and therefore with below-zero copyright).”

---

## About libraries

The Helsinki City Library in Finland was the first public library to create its own website in February 1994. According to the report “Internet and the Library Sphere: Further Progress for European Libraries”, a report published online by the European Commission, 1,000 public libraries from 26 European countries had their own websites in December 1998. The websites ranged from one web page giving the library’s physical address and opening hours to full websites with access to OPACs (online public access catalogues) and other services. The leading countries were Finland (247 libraries), Sweden (132 libraries), the United Kingdom (112 libraries), Denmark (107 libraries), Germany (102 libraries), Netherlands (72 libraries), Lithuania (51 libraries), Spain (56 libraries) and Norway (45 libraries). Newcomers were the Czech Republic (29 libraries) and Portugal (3 libraries). Russia provided a web page with a list of 26 public reference libraries.

Gabriel, which stands for “Gateway and Bridge to Europe’s National Libraries”, was created in January 1997 by the Conference of European National Librarians (CENL) as a trilingual (English, French, German) website. As explained on the website: “Gabriel also recalls Gabriel Naudé, whose ‘*Advis pour dresser une bibliothèque*’ [‘Instructions Concerning Erecting a Library’] (Paris, 1627) is one of the earliest theoretical works about libraries in any European language, and provides a blueprint for the great modern research library. The name Gabriel is common to many European languages and is derived from the Old Testament, where Gabriel appears as one of the archangels or heavenly messengers. He also appears in a similar role in the New Testament and the Quran.”

One year later, Gabriel offered links to the internet services of 38 participating national libraries (Albania, Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russia, San Marino, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, Vatican City). The internet services were OPACs, national bibliographies, national union catalogues, and indexes for periodicals. A specific web page listed common European projects.

Much later, in summer 2005, Gabriel merged with the website of the European Library (created by CNL in January 2004) to offer a web portal for 43 national libraries. Europeana, the European digital library, was launched in November 2008 with two million documents. It offered 10 million documents in September 2010.

---

## About librarians

In “Books in My Life”, a book published by the Library of Congress in 1985, Robert Downs wrote: “My lifelong love affair with books and reading continues unaffected by automation, computers, and all other forms of the twentieth-century gadgetry.” But automation, computers and the internet eased the work of many librarians, for example Peter Raggett at the Organisation for Economic Cooperation and Development (OECD) and Bruno Didier at the Pasteur Institute. Their main task was to help their patrons find the information they needed at a time when web search engines were less accurate.

The OECD Library was among the first ones in Europe to set up an extensive intranet for its staff. The core of OECD’s original members had expanded from Europe and North America to include Japan, Australia, New Zealand, Finland, Mexico, the Czech Republic, Hungary, Poland and South Korea. The library offered 60,000 monographs and 2,500 journals in 1999, as well as microfilms and CD-ROMs, and subscriptions to databases such as Dialog, Lexis-Nexis and UnCover.

Peter Raggett, the library head, first worked in government libraries in the United Kingdom before joining the OECD in 1994. He wrote in August 1999 in an email interview: “At the OECD Library we have collected together several hundred websites and have put links to them on the OECD intranet. They are sorted by subject and each site has a short annotation giving some information about it. The researcher can then see if it is possible that the site contains the desired



information. This is adding value to the site references and in this way the Central Library has built up a virtual reference desk on the OECD network. As well as the annotated links, this virtual reference desk contains pages of references to articles, monographs and websites relevant to several projects currently being researched at the OECD, network access to CD-ROMs, and a monthly list of new acquisitions. The Library catalogue will soon be available for searching on the intranet. The reference staff at the OECD Library uses the internet for a good deal of their work. Often an academic working paper will be on the web and will be available for full-text downloading. We are currently investigating supplementing our subscriptions to some of our periodicals with access to the electronic versions on the internet.”

“The internet has provided researchers with a vast database of information. The problem for them is to find what they are seeking. Never has the information overload been so obvious as when one tries to find information on a topic by searching the internet. When one uses a search engine like Lycos or AltaVista or a directory like Yahoo!, it soon becomes clear that it can be very difficult to find valuable sites on a given topic. These search mechanisms work well if one is searching for something very precise, such as information on a person who has an unusual name, but they produce a confusing number of references if one is searching for a topic which can be quite broad. Try and search the web for Russia \*and\* transport to find statistics on the use of trains, planes and buses in Russia. The first references you will find are freight-forwarding firms that have business connections with Russia.”

“The internet is impinging on many people’s lives, and information managers are the best people to help researchers around the labyrinth. The internet is just in its infancy and we are all going to be witnesses to its growth and refinement. Information managers have a large role to play in searching and arranging the information on the internet. I expect that there will be an expansion in internet use for education and research. This means that libraries will have to create virtual libraries where students can follow a course offered by an institution at the other side of the world. Personally, I see myself becoming more and more a virtual librarian. My clients may not meet me face-to-face but instead will contact me by email, telephone or fax, and I will do the research and send them the results electronically.”

Another experience is that of Bruno Didier, webmaster of the Pasteur Institute Library. The Pasteur Institutes are observatories for studying infectious and parasite-borne diseases (malaria, tuberculosis, AIDS, yellow fever, dengue, poliomyelitis, and others). Bruno Didier explained in August 1999: “The main aim of the Pasteur Institute Library’s website is to serve the Institute itself and its associated bodies. It supports applications that have become essential in such a big organisation: bibliographic databases, cataloguing, ordering of documents, and of course access to online periodicals (more than 100 titles). It is a window for our different departments, at the Pasteur Institute here in Paris but also elsewhere in France and abroad. It is very useful to exchange information with the Pasteur Institutes worldwide. The website has existed in its present form since 1996, and the number of users is steadily increasing.”

“We see a change in our relationship with both the information and the users. We increasingly become mediators, and perhaps to a lesser extent curators. My present activity is typical of this new situation: I provide quick access to information, I create effective means of communication, and I also train people to use these new tools. I think that, in the future, our work will be based on cooperation and on the use of common resources. It is an old wish for librarians, but it is the first time we have the means to realise it.”

---

## About digital libraries

With a digital library, librarians could finally fulfil two goals that used to be in contradiction: preservation (on shelves) and communication (on the internet). People could now leaf through digital facsimiles, and access the original works only when necessary. On the one hand, physical books were taken out of their shelves only once to be scanned. On the other hand, digitised books could easily be accessed anywhere at any time, without the need to go to the library and to struggle through a lengthy process to access the original works, because of reduced opening hours, forms to fill out, safety concerns for rare books, and shortage of staff. Some researchers probably remember the unfailing patience and determination needed to access a rare book or a manuscript.

Brian Lang, chief executive of the British Library, explained in 1998 on its new website: “We do not envisage an exclusively digital library. We are aware that some people feel that digital materials will predominate in libraries of the future. Others anticipate that the impact will be slight. In the context of the British Library, printed books, manuscripts, maps, music, sound recordings and all the other existing materials in the collection will always retain their central importance, and we are committed to continuing to provide, and to improve, access to these in our reading rooms. The importance of digital materials will, however, increase. We recognise that network infrastructure is at present most strongly developed in the higher education sector, but there are signs that similar facilities will also be available elsewhere, particularly in the industrial and commercial sector, and for public libraries. Our vision of network access encompasses all these. The development of the Digital Library [in February 1999] will enable the British Library to embrace the digital information age. Digital technology will be used to preserve and extend the Library’s unparalleled collection. Access to the collection will become boundless with users from all over the world, at any time, having simple, fast access to digitised materials using computer networks, particularly the internet.”

The French National Library launched its digital library Gallica in October 1997 with the digitised versions of 2,500 books from the 19th century relating to French history, life and culture. When interviewed by Jérôme Strazzulla, journalist for the daily *Le Figaro*, for its 3 June 1998 issue, Jean-Pierre Angremy, president of the library, stated: “We cannot, we will not be able to digitise everything. In the long term, a digital library will only be one element of the whole library.” The books later ranged from the Middle Ages to the early 20th century. They were first available as bulky image files before being supplemented with easier-to-use text files.

Many image collections went online, for example 15,000 historical and contemporary images from the National Library of Australia’s Pictorial Collection, including paintings, drawings, rare prints and photographs.

Digital collections became global, with Google Print (renamed Google Books) launched in October 2004, the Open Content Alliance (OCA) launched in October 2005, Europeana launched in November 2008, and the Digital Public Library of America (DPLA) launched in April 2013.

The Open Content Alliance (OCA) started with an idea from the Internet Archive, founded in April 1996 by Brewster Kahle in San Francisco. According to its website in 2007, OCA was “a collaborative effort of a group of cultural, technology, nonprofit, and governmental organisations from around the world that helps build a permanent archive of multilingual digitised text and multimedia material. An archive of contributed material is available on the Internet Archive website and through Yahoo! and other search engines and sites. The OCA encourages access to and reuse of collections in the archive, while respecting the content owners and contributors.”

The project aimed at digitising public domain books and other content worldwide. Unlike Google Books, the Open Content Alliance only offers public domain works, except when the copyright holder has expressly given permission, and the books can be accessed on any web search engine. The first contributors were the University of California, the University of Toronto, the European Archive, the National Archives of the United Kingdom, O'Reilly Media and Prelinger Archives. The collection included 100,000 e-books in December 2006, 200,000 e-books in May 2007, one million e-books in December 2008, and two million e-books in March 2010.

---

## About treasures of the past

Libraries began digitising their treasures for the world to enjoy. For example, the British Library digitised Beowulf, the earliest known narrative poem in English. The library holds the only known manuscript of Beowulf, dated around the year 1000. The poem itself is much older than the manuscript — it might have been written around 750. The manuscript was badly damaged by fire in 1731. Early 18th-century transcripts mentioned hundreds of words and characters which crumbled away along the charred edges over the years. To halt this process, each leaf was mounted on a paper frame in 1845.

Researchers around the world have regularly requested access to the manuscript. Taking Beowulf out of its display case for study not only raised conservation issues, but it also made it unavailable for the many visitors of the British Library expecting to see Beowulf on display. Digitising the manuscript offered a solution to these problems, and new opportunities for researchers and readers worldwide.

Brian Lang, chief executive of the British Library, explained on its website in 1998: “The Beowulf manuscript is a unique treasure and imposes on the Library a responsibility to scholars throughout the world. Digital photography offered for the first time the possibility of recording text concealed by early repairs, and a less expensive and safer way of recording readings under special light conditions. It also offers the prospect of using image enhancement technology to settle doubtful readings in the text. Network technology has facilitated direct collaboration with American scholars and makes it possible for scholars around the world to share in these discoveries. Curatorial and computing staff learned a great deal which will inform any future programmes of digitisation and network service provision the Library may undertake, and our publishing department is considering the publication of an electronic scholarly edition of Beowulf. This work

has not only advanced scholarship; it has also captured the imagination of a wider public, engaging people (through press reports and the availability over computer networks of selected images and text) in the appreciation of one of the primary artefacts of our shared cultural heritage.”

Other digitised treasures of the British Library were already available online, for example Magna Carta, a charter agreed by King John of England in 1215 with its Great Seal, and considered the first constitutional text ever; the Lindisfarne Gospels, an illuminated manuscript gospel book produced around the year 700; the Diamond Sutra, dated 868 and “the earliest complete survival of a dated printed book”; the Sforza Hours, a richly illuminated book of hours dated 1490-1520; the Codex Arundel, a bound collection of notes written by Leonardo Da Vinci between 1480 and 1518; and the Tyndale New Testament, the first English New Testament to be printed in 1526.

The original Gutenberg Bible was available online in November 2000. Gutenberg printed his first Bibles in 1454 or 1455 in Mainz, Germany, perhaps printing 180 copies, with 48 copies still available in 2000, and three copies (two full copies and one partial copy) belonging to the British Library. The two full copies — a little different from each other — were digitised in March 2000 by Japanese experts from the Keio University of Tokyo and NTT (Nippon Telegraph and Telephone Communications). The images were then processed to offer a full digitised version on the web a few months later.

Many libraries digitised their own treasures, for example the Bielefeld University Library in Germany. Michael Behrens, in charge of the digital library project, wrote in September 1998 in an email interview: “We started digitising rare prints from our own library, and some rare prints that were sent in via library loan in November 1996. In that first phase of our attempts at digitisation, starting in November 1996 and ending in June 1997, 38 rare prints were scanned as image files and made available on the web. In the same period, there were also a few digital materials prepared as accompanying materials for lectures held at the university (image files for excerpts from printed works). These are, for copyright reasons, not available outside the campus. The next step, which is just being completed, is the digitisation of the *Berlinische Monatsschrift*, a German periodical from the Enlightenment, comprising 58 volumes, with 2,574 articles and 30,626 pages. A larger project to digitise German periodicals from the 18th and early 19th centuries (around one million pages) is planned for soon. These periodicals belong to our library’s collection and to other libraries elsewhere. The digitisation project would be coordinated here, and some of the technical work would be done here too.”

The experimental database of the first volume (1751) of the *Encyclopédie* by Diderot and d’Alembert was available online in 1998 on the website of ARTFL (American and French Research on the Treasury of the French Language), a joint project of the University of Chicago and the French National Center for Scientific Research (CNRS). The database of the first volume was the first step towards a full online edition of the 17 volumes of text (with 18,000 pages and 21.7 million words) and 11 volumes of plates of the *Encyclopédie* (1st edition, 1751-72), with 72,000 articles written by 140 contributors (Diderot, d’Alembert, Voltaire, Rousseau, Marmontel, d’Holbach, Turgot, and others). Designed to collect and disseminate the entire knowledge of that time, the *Encyclopédie* was a reflection of the intellectual and social currents of the Enlightenment, and contributed to disseminate novel ideas that would inspire the French Revolution in 1789.

---

## About library catalogues

The MARC (Machine-Readable Cataloguing) format was created by the International Federation of Library Associations (IFLA) in the early 1970s. As explained in “UNIMARC: An Introduction” in 1999, it was “a short and convenient term for assigning labels to each part of a catalogue record so that it can be handled by computers. While the MARC format was primarily designed to serve the needs of libraries, the concept has since been embraced by the wider information community as a convenient way of storing and exchanging bibliographic data.”

Several versions of MARC emerged over the years because of different national cataloguing practices. With twenty MARC formats (INTERMARC, USMARC, UKMARC, CAN/MARC, etc.), differences in data content meant extensive editing before and after records were exchanged. UNIMARC (Universal Machine-Readable Cataloguing) was published in 1977 by the IFLA as a common bibliographic format, with an updated version in 1980, and a handbook in 1983 and 1987. Cataloguers could now process records created in any MARC format. Records in one MARC format were first converted into UNIMARC before being converted into another MARC format. Each national bibliographic agency needed to write only two programs — one to convert into UNIMARC and one to convert from UNIMARC — instead of the several programs that were needed until then for the conversion of each MARC format.

UNIMARC was also promoted as a format on its own, and was adopted by several national bibliographic agencies as their in-house format. A Permanent UNIMARC Committee was created in 1991 to monitor the development of UNIMARC. The European Commission officially promoted UNIMARC in 1996 as its preferred bibliographic format for libraries in the European Union. The British Library (using UKMARC), the Library of Congress (using USMARC) and the National Library of Canada (using CAN/MARC) harmonised their national MARC formats and published their common format MARC 21 in 1999.

Union catalogues thrived. The idea behind a union catalogue was to avoid cataloguing the same document by many cataloguers worldwide. When cataloguers of a member library processed a new document, they first searched the union catalogue. If the record was available, they imported it into their own catalogue, and added the local data. If the record was not available, they created it in their own catalogue and exported it into the union catalogue, for it to be available to all cataloguers of member libraries. The two main global union catalogues (paid subscription) were the RLG Union Catalog and OCLC's WorldCat.

Created in 1980 by the Research Libraries Group (RLG), the RLG Union Catalog was first known under the name of RLIN (Research Libraries Information Network). RLIN included the records of 88 million documents held in libraries belonging to RLG member institutions in 1998. There were mainly research and specialised libraries, for example law, technical and corporate libraries. RLIN included several records per document (unlike OCLC's WorldCat with one record per document).

RLG member institutions were for example the Library of Congress, the National Library of Medicine, the U.S. Government Printing Office, CONSER (Conversion of Serials Project), the British Library, the British National Bibliography, and the National Union Catalogue of Manuscript collections. RLIN provided access to special resources, for example the United Nations' DOCFILE and CATFILE records, the Rigler Deutsch Index to pre-1950 commercial sound recordings, the catalogue of the machine-readable data files of the French literary works in the ARTFL Database, the statistical data collected by ICPSR (Inter-university Consortium for Political and Social Research) at the University of Michigan, and the catalogue of the archival and manuscript collections of research libraries, museums, state archives, and historical societies in North America.

RLIN also hosted the English Short Title Catalogue (ESTC), with records of the letterpress materials printed in the United Kingdom or its dependencies in any language, from the beginnings of print to 1800, as well as materials printed in English worldwide. Produced by the ESTC editorial offices at the University of California (Riverside) and at the British Library, with the help of the American Antiquarian Society and 1,600 libraries worldwide, ESTC was updated daily as a comprehensive bibliography of the hand-press era and a census of surviving copies. Materials ranged from Shakespeare's works and Greek New Testaments to anonymous ballads, songs, advertisements, and other ephemera.

RLIN was renamed RLG Union Catalogue in 2003. RedLightGreen, its free experimental web version, was launched in fall 2003, followed by a full version in spring 2004. RedLightGreen closed its site in November 2006 when the RLG team joined OCLC, with a link to OCLC's WorldCat.

OCLC (Online Computer Library Center) was created in 1967 as a non-profit organisation dedicated to "furthering access to the world's information while reducing information costs". The OCLC Online Union Catalog (paid subscription) began as a regional computer catalogue for the 54 college and university libraries in the State of Ohio. It became a national union catalogue for the whole country before becoming a worldwide union catalogue with an international network of libraries. Renamed WorldCat, the catalogue included 38 million records in 400 languages in 1998, with 2 million records added annually. OCLC's website was available in six languages (English, Chinese, French, German, Portuguese, Spanish). Libraries inside the United States received OCLC services through their OCLC-affiliated regional networks. Libraries outside the United States received OCLC services through OCLC Asia Pacific, OCLC Canada, OCLC Europe, OCLC Latin America and the Caribbean, or via international distributors.

WorldCat included 61 million bibliographic records in 400 languages in 2005, from 9,000 member libraries in 112 countries. It included 73 million bibliographic records in 2006, with links to one billion documents available in these libraries. WorldCat launched the beta version of its new website [worldcat.org](http://worldcat.org) in August 2006, for member libraries to provide free access to their catalogues, and free or paid access to their electronic resources (books, audio books, abstracts, full-text articles, photos, music CDs, videos). 1.5 billion documents could be located and/or accessed using WorldCat in April 2010.

---

## About linguistic resources

The first online translation dictionaries for travellers were offered by Travlang. Michael C. Martin, a physics student in New York, created the web page “Foreign Languages for Travelers” in 1994 on the website of his university, and transferred it on his new website Travlang one year later. He moved to California to work as a researcher in experimental physics at the Lawrence Berkeley National Laboratory, and expanded Travlang on his free time.

Michael C. Martin wrote in August 1998 in an email interview: “I think the web is an ideal place to bring different cultures and people together, and that includes being multilingual. Our Travlang site is so popular because of this, and people desire to feel in touch with other parts of the world. The internet is really a great tool for communicating with people you wouldn’t have the opportunity to interact with otherwise. I truly enjoy the global collaboration that has made our ‘Foreign Languages for Travelers’ pages possible.”

It offered online resources to learn 60 languages in 1998, while the web page “Translating Dictionaries” gave access to free online dictionaries in 16 languages (Afrikaans, Czech, Danish, Dutch, Esperanto, Finnish, French, Frisian, German, Hungarian, Italian, Latin, Norwegian, Portuguese, Spanish, Swedish). The website also offered links to translation services, language schools, plurilingual bookstores and travel services. People could book their hotel, car or plane ticket, check exchange rates, and browse a directory of language and travel websites. Michael C. Martin sold Travlang in February 1999.

Tyler Chambers, a software developer in Boston, Massachusetts, launched two projects on his free time, the Human-Languages Page in 1994 and the Internet Dictionary Project in 1995. He explained in September 1998 in an email interview: “1994 was the year I was really introduced to the web, which was a little while after its christening but long before it was mainstream. That was also the year I began my first multilingual web project, and there was already a significant number of language-related resources online. This was back before Netscape even existed — Mosaic was almost the only web browser, and web pages were little more than hyperlinked text documents.”

The Human-Languages Page (H-LP) was a directory of online linguistic resources, with links to 1,800 resources in 100 languages in 1998. It merged with the Languages Catalog, a section of the WWW Virtual Library, to become iLoveLanguages in spring 2001. iLoveLanguages provided an index of 2,000 linguistic resources in 100 languages in September 2003.

The Internet Dictionary Project (IDP) was a collaborative project to create free online dictionaries from English to six other languages (French, German, Italian, Latin, Portuguese, Spanish). As explained on its website: “The Internet Dictionary Project began in 1995 in an effort to provide a noticeably lacking resource to the internet community and to computing in general — free translating dictionaries. Not only is it helpful to the online community to have access to dictionary searches at their fingertips via the World Wide Web, it also sponsors the growth of computer software which can benefit from such dictionaries — from translating programs to spelling-checkers to language-education guides and more. By facilitating the creation of these dictionaries online by thousands of anonymous volunteers all over the internet, and by providing the results

free-of-charge to anyone, the Internet Dictionary Project hopes to leave its mark on the internet and to inspire others to create projects which will benefit more than a corporation's gross income. (...) This site allows individuals from all over the world to visit and assist in the translation of English words into other languages. The resulting lists of English words and their translated counterparts are then made available through this site to anyone, with no restrictions on their use."

Tyler Chambers wrote in the same email interview: "While I'm not multilingual, nor even bilingual, myself, I see an importance to language and multilingualism that I see in very few other areas. The internet has allowed me to reach millions of people and help them find what they're looking for, something I'm glad to do. Overall, I think that the web has been great for language awareness and cultural issues — where else can you randomly browse for 20 minutes and run across three or more different languages with information you might potentially want to know? Communications media make the world smaller by bringing people closer together; I think that the web is the first (of mail, telegraph, telephone, radio, TV) to really cross national and cultural borders for the average person. I think that the future of the internet is even more multilingualism and cross-cultural exploration and understanding than we've already seen. But the internet will only be the medium by which this information is carried; like the paper on which a book is written, the internet itself adds very little to the content of information, but adds tremendously to its value in its ability to communicate that information."

"To say that the internet is spurring multilingualism is a bit of a misconception, in my opinion — it is communication that is spurring multilingualism and cross-cultural exchange, the internet is only the latest mode of communication which has made its way down to the (more-or-less) common person. Language will become even more important than it already is when the entire planet can communicate with everyone else (via the web, chat, games, email, and whatever future applications haven't even been invented yet), but I don't know if this will lead to stronger language ties, or a consolidation of languages until only a few, or even just one remain. One thing I think is certain is that the internet will forever be a record of our diversity, including language diversity, even if that diversity fades away. And that's one of the things I love about the internet — it's a global model of the saying 'it's not really gone as long as someone remembers it. And people do remember. (...) As browsers and users mature, I don't think there will be any currently spoken language that won't have a niche on the web, from Native American languages to Middle Eastern dialects, as well as a plethora of 'dead' languages that will have a chance to find a new audience with scholars and others alike online."

Tyler Chambers ran out of time to maintain the Internet Dictionary Project, and removed the ability to update the dictionaries in January 2007. Users could still search the dictionaries on the website and download the archived files.

Logos, an Italian translation company, decided in December 1997 to make its professional tools freely available on the web for all its translators and for the general public. These professional tools were: (1) the Logos Dictionary, a multilingual dictionary with 7.5 billion words (in fall 1998); (2) the Logos Wordtheque, a multilingual library with 328 million words extracted from translated novels, technical manuals and other texts, that could be searched by language, word, author or title; (3) the Logos Linguistic Resources, a database of 500 glossaries; and (4) the Logos Universal Conjugator, a database of verbs in 17 languages.



Founded by Rodrigo Vergara in 1979 in Modena, Italy, Logos employed 200 in-house translators and 2,500 free-lance translators who processed around 200 texts per day in 1997. When interviewed by journalist Annie Kahn for her article “Les mots pour le dire” (The words to tell it) published by the French daily Le Monde on 7 December 1997, Rodrigo Vergara explained: “We wanted all our translators to have access to the same translation tools. So we made them available on the internet, and while we were at it we decided to make the site open to the public. This made us extremely popular, and also gave us a lot of exposure. This move has in fact attracted many customers, and it has allowed us to widen our network of translators in the wake of this initiative.”

Annie Kahn wrote in the same article: “The Logos site is much more than a mere dictionary or a collection of links to other online dictionaries. The cornerstone is the document search, which processes a corpus of literary texts available free of charge on the web. If you search for the definition or the translation of a word (‘didactics’, for example), you get not only a reply but also a quote from one of the literary works using the word (in this case, an essay by Voltaire). All it takes is a click of the mouse to access the whole text and even order the book, including in foreign translations, thanks to a partnership with the well-known online bookstore Amazon.com. However, if there is no text using the word in the database, the software acts as a search engine suggesting other web sources. For some words, you can even hear the pronunciation. If there is no available translation, the software sends a request to the general public. Everyone can make suggestions, and the translators at Logos check these suggestions to decide if they can validate them and include them in the database.”

Ten years later, in 2007, the Logos Library (formerly Wordtheque) included 710 billion words, Linguistic Resources (no change of name) included 1,215 glossaries, and Conjugation of Verbs (formerly Universal Conjugator) included verbs in 36 languages.

---

## About dictionaries

The first online reference dictionaries were based on their printed counterparts. Created in 1996, the website “Merriam-Webster Online: The Language Center” gave free access to the digitised edition of several print publications: Webster Dictionary, Webster Thesaurus, Webster’s Third (a lexical landmark), Guide to International Business Communications, Vocabulary Builder (with interactive vocabulary quizzes), and Barnhart Dictionary Companion (‘hot’ new words). The website also tracked down definitions, spellings, pronunciations, synonyms, vocabulary exercises, and other key facts about words and language. Created in 1997, the Dictionnaire Universel Francophone en Ligne (Universal French-Language Online Dictionary) was the free online version of the printed dictionary published by Hachette. Created in March 2000, the online version (for a subscription fee) of the 20-volume Oxford English Dictionary (OED) was updated quarterly with 1,000 new or revised entries.

The French-English GDT (Grand Dictionnaire Terminologique – Large Terminology Dictionary) was a free online dictionary launched in September 2000 with 3 million terms relating to industry, science and commerce. The GDT was designed directly for the web by the Quebec Office of the French Language (OQLF), with a database created and maintained by Semantix. The GDT was

used by 1.3 million people during the first month, with peaks of 60,000 visits per day, which certainly contributed to better translations. The database was then maintained by Convera Canada, with 3.5 million visits per month in February 2003. A new version of the GDT went online in March 2003, with the database maintained by OQLF itself, and the addition of Latin as a third language.

There were also dictionary portals. Robert Beard, a professor at Bucknell University in Lewisburg, Pennsylvania, created A Web of Online Dictionaries (WOD) in 1995 as a directory of online dictionaries and linguistic resources (thesauri, vocabularies, glossaries, grammars, textbooks), with 800 dictionaries in 150 languages in September 1998. The section Web of Linguistic Fun made linguistics appealing for non-specialists.

Robert Beard wrote in September 1998 in an email interview: “There was an initial fear that the web posed a threat to multilingualism on the web, since HTML and other programming languages are based on English and since there are simply more websites in English than any other language. However, my website indicates that multilingualism is very much alive and the web may, in fact, serve as a vehicle for preserving many endangered languages. Moreover, the new attention paid by browser developers to the different languages of the world will encourage even more websites in different languages.”

He added in January 2000: “A Web of Online Dictionaries (WOD) is now part of yourDictionary.com. The new website is an index of 1,200+ dictionaries in more than 200 languages. Besides the WOD, the new website includes a word-of-the-day feature, word games, a language chat room, the old Web of Online Grammars (now expanded to include additional language resources), the Web of Linguistic Fun, multilingual dictionaries, specialised English dictionaries, thesauri and other vocabulary aids, language identifiers and guessers, and dictionary indices. yourDictionary.com will hopefully be the premier language portal and the largest language resource site on the web. It is now actively acquiring dictionaries and grammars of all languages with a particular focus on endangered languages. It is overseen by a blue ribbon panel of linguistic experts from all over the world.”

“yourDictionary.com has lots of new ideas. We will have language chat rooms and bulletin boards. There will be language games designed to entertain and teach fundamentals of linguistics. The Linguistic Fun page will become an online journal for short, interesting, yes, even entertaining, pieces on language that are based on sound linguistics by experts from all over the world. We plan to work with the Endangered Language Fund in the U.S. and Britain to raise money for the Foundation’s work and publish the results on our site in the Endangered Language Repository. Languages that are endangered are primarily languages without writing systems at all (only 1/3 of the world’s 6,000+ languages have writing systems). I still do not see the web contributing to the loss of language identity and still suspect it may, in the long run, contribute to strengthening it. More and more Native Americans, for example, are contacting linguists, asking them to write grammars of their language and help them put up dictionaries. For these people, the web is an affordable boon for cultural expression.”

Michael Kellogg created WordReference.com in 1999, and explained on its website: “I started this site as an effort to provide free online bilingual dictionaries and tools to the world. The site has grown gradually ever since to become one of the most used online dictionaries, and the top online dictionary for its language pairs of English-Spanish, English-French, English-Italian, Spanish-

French, and Spanish-Portuguese. It is consistently ranked in the top 500 most visited websites in the world. I am proud of my history of innovation with dictionaries on the internet. Many of the features such as being able to click any word in a dictionary entry were first implemented by me.”

“The internet has done an incredible job of bringing the world together in the last few years. Of course, one of the greatest barriers has been language. Much of the content is in English and many, many users are reading English-language web pages as a second language. I know from my own experience with Spanish-language websites that many readers probably understand much of what they are reading, but not every single word. Today, I have three main goals with my website. First, continue to create free online bilingual dictionaries from English to many other languages. I strive to offer translations for \*all\* English words, terms, idioms, sayings, etc. Second, provide the world’s best language forums; and third, continue to innovate to produce the best website and tools for the world.”

In 2010, WordReference offered an English monolingual dictionary, and dictionaries from English to other languages (Arabic, Chinese, Czech, Greek, Japanese, Korean, Polish, Portuguese, Romanian, Turkish), and vice versa. It offered a Spanish monolingual dictionary, a Spanish dictionary of synonyms, a Spanish-French dictionary and a Spanish-Portuguese dictionary. There was a monolingual dictionary for German, and another one for Russian. Conjugation tables were available for French, Italian and Spanish. WordReference Mini was a miniature version of the site to be embedded into other sites, for example sites teaching languages online. There was a mobile device version for dictionaries from English to French, English to Italian and English to Spanish, and vice versa, with other language pairs planned for later.

---

## About encyclopedias

The first reference encyclopedias were available online in late 1999. They were based on their printed counterparts before been designed directly for the web.

Britannica.com was created in December 1999 as the digital equivalent of the 32 volumes of the Encyclopaedia Britannica’s 15th edition. Britannica.com was available for free, as a complement to the printed and CD-ROM editions for sale. It offered links to articles from 70 magazines, to websites, to books, etc., all of them searchable through a single search engine. Britannica.com joined the top 100 websites in the world in September 2000. It switched from being free to being accessed for a monthly or yearly subscription fee in July 2001. It opened its website to external contributors in 2009, with registration required to write and edit articles.

WebEncyclo was the first main French-language online encyclopedia in December 1999, with a free content based on the printed encyclopedia published by Éditions Atlas. WebEncyclo was searchable by keyword, topic and media (maps, links, photos, illustrations). A call for papers invited specialists to become external contributors and submit their articles in the section WebEncyclo Contributif. Later on, a free registration was required to use the online encyclopedia.

The website of the printed Encyclopaedia Universalis — the French-language sister of Encyclopaedia Britannica — was also created in December 1999. It included 28,000 articles by 4,000 contributors, available for an annual subscription fee, with a number of articles available for free.

Two years after the creation of the 20-volume Oxford English Dictionary's online version, the Oxford University Press (OUP) created in March 2002 Oxford Reference Online (ORO), a comprehensive encyclopedia (paid subscription) designed directly for the web. According to its publisher, its 60,000 web pages and one million entries were the equivalent of 100 printed encyclopedias.

Then came the free global collaborative online encyclopedias, that would change the way we learn and the way we share our knowledge.

Wikipedia was launched in January 2001 by Jimmy Wales and Larry Sanger (Larry Sanger resigned later on). One year later, it offered 20,000 articles in 18 languages, that could be freely reused under a GFDL license (and a Creative Commons license later on). Wikipedia quickly became the largest reference website, with thousands of people contributing worldwide. It offered 1.3 million articles by 13,000 contributors in 100 languages in 2004. It was in the top ten websites worldwide in 2006, with 6 million articles in 250 languages. It offered 7 million articles in 192 languages in May 2007, including 1.8 million articles in English, 589,000 articles in German, 500,000 articles in French, 260,000 articles in Portuguese, and 236,000 articles in Spanish. It was in the top five websites worldwide in 2008. It offered 14 million articles in 272 languages in September 2010, including 3.4 million articles in English, 1.1 million articles in German, and 1 million articles in French. Wikipedia celebrated its tenth anniversary in January 2011 with 17 million articles in 270 languages, et 400 million individual visits per month for all the Wikimedia websites (Wikipedia, Wiktionary, Wikibooks, Wikiquote, Wikisource, Wikipedia Commons, Wikispecies, Wikinews, Wikiversity).

Citizendium, which stands for “The Citizen’s Compendium”, was created in March 2007 by Larry Sanger as a pilot project to build a free global collaborative online encyclopedia led by experts. Larry Sanger, who co-founded Wikipedia in January 2001, had left the team over policy and content quality issues, and the use of anonymous pseudonyms for contributors. Citizendium wanted to combine “public participation with gentle expert guidance” in a project that was experts-led, but not experts-only. Contributors used their own names, and were guided by expert editors. Constables made sure that the rules are respected.

As explained by Larry Sanger in his essay “Toward a New Compendium of Knowledge” on the website of Citizendium: “Editors will be able to make content decisions in their areas of specialisation, but otherwise working shoulder-to-shoulder with ordinary authors.” There were 1,100 articles from 820 authors and 180 editors in March 2007, 11,800 articles in August 2009, and 15,000 articles in September 2010. Citizendium also wanted to act as a prototype for large scale knowledge-building projects that would deliver reference scholarly and educational content.

The Encyclopedia of Life was created in May 2007 as a global scientific effort to document all known species of animals and plants. There were 1.8 million species, including endangered species, and probably 6 to 8 million species yet to be discovered and catalogued. This collaborative

effort was led by several major institutions (Field Museum of Natural History, Harvard University, Marine Biological Laboratory, Missouri Botanical Garden, Smithsonian Institution, Biodiversity Heritage Library).

The encyclopedia's honorary chair was Edward Wilson, professor emeritus at Harvard University, who, in an essay dated 2002, was the first to express the wish for such an encyclopedia as a single portal for millions of documents scattered online and offline. Technology improvements made it possible five years later, with content aggregators, mash-up, wikis and large-scale content management to process texts, photos, maps, sound and videos. The first pages of the encyclopedia were available in mid-2008, with one web page for each species. The English version was expected to be translated in several languages by partner organisations.

---

## About journals

The Public Library of Science (PLOS), founded in October 2000, first advocated for scientific journals to be freely available in online archives. PLOS created a non-profit scientific and medical publishing venture in early 2003 to provide scientists and physicians with free high-quality peer-reviewed online journals in which they could publish their work. The journals were PLOS Biology (2003), PLOS Medicine (2004), PLOS Genetics (2005), PLOS Computational Biology (2005), PLOS Pathogens (2005), PLOS Clinical Trials (2006), and PLOS Neglected Tropical Diseases (2007), the first scientific journal on this topic. PLOS also created PLOS One (2006), a journal covering primary research for any discipline in science and medicine. All PLOS articles are freely available online, on the websites of PLOS and in PubMed Central, the public archive of the National Library of Medicine. The articles can be freely redistributed and reused under a Creative Commons license, including for translations, as long as the author(s) and source are cited.

PLOS received financial support from several foundations while developing a viable economic model from fees paid by published authors, advertising, sponsorship, and paid activities organised for PLOS members. Three years after their creation, PLOS Biology and PLOS Medicine had the same reputation of excellence as the leading fee-based scientific journals *Nature*, *Science* and *The New England Journal of Medicine*. PLOS' goal was also to encourage other publishers to adopt an open access model, or to convert their existing journals to an open access model.

What is open access? The Budapest Open Access Initiative (BOAI) was signed in February 2002 as the founding text of the open access movement, available in several languages. "By 'open access' to [research] literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

The Directory of Open Access Journals (DOAJ) was created in 2003 as a directory of open access scholarly and scientific journals in any field and any language, with 9,800 journals on 29 December 2013, 10,068 journals on 15 November 2014 (with half of them searchable at article level), 10,224 journals on 18 February 2015, and 10,459 journals from 134 countries on 20 June 2015. With a new website launched in December 2013, DOAJ became the authoritative source for open access journals who were either peer reviewed or using a quality control system. Their inclusion in DOAJ increased their visibility and impact. DOAJ's goal was also to encourage best practices among open access publishers.

The open access movement also encouraged open archives. Major universities created their open archive, for example DASH (Digital Access to Scholarship at Harvard) created by Harvard and DSpace@MIT created by the Massachusetts Institute of Technology (MIT). DASH was created in September 2009 as an open access repository for members of the Harvard community, for their work to be freely accessed worldwide, with 4.1 million downloads in November 2014, 4.7 million downloads in February 2015, and 5.5 million downloads in June 2015. DSpace@MIT is an open access repository for peer-reviewed articles, technical reports, working papers, theses and more, with 60,000+ items in December 2013 and one million end-user downloads per month, and with 70,000+ items in February 2015.

---

## About resources for teaching

When interviewed by the French daily *Libération* in January 1998, Vinton Cerf, often named the father of the internet as the co-inventor of internet protocols, explained that the internet was doing two things. First, it provided some information, like a book. Second, information was connected with other information while, in a book, information stayed isolated.

More and more computers connected to the internet were available in schools and at home in the mid-1990s, and teachers began exploring new ways of teaching. Dale Spender, an Australian scholar and teacher, gave a lecture on “Creativity and the Computer Education Industry” during a conference held by the International Federation of Information Processing (IFIP) in September 1996.

“Throughout print culture, information has been contained in books — and this has helped to shape our notion of information. For the information in books stays the same — it endures. And this has encouraged us to think of information as stable — as a body of knowledge which can be acquired, taught, passed on, memorised, and tested of course. The very nature of print itself has fostered a sense of truth; truth too is something which stays the same, which endures. And there is no doubt that this stability, this orderliness, has been a major contributor to the huge successes of the industrial age and the scientific revolution. (...) But the digital revolution changes all this. Suddenly it is not the oldest information — the longest lasting information that is the most reliable and useful. It is the very latest information that we now put the most faith in — and which we will pay the most for.”

“Education will be about participating in the production of the latest information. This is why education will have to be ongoing throughout life and work. Every day there will be something new that we will all have to learn. To keep up. To be in the know. To do our jobs. To be members of the digital community. And far from teaching a body of knowledge that will last for life, the new generation of information professionals will be required to search out, add to, critique, ‘play with’, and daily update information, and to make available the constant changes that are occurring.”

Robert Beard, professor at Bucknell University in Lewisburg, Pennsylvania, wrote in September 1998 in an email interview: “The web represents a plethora of new resources produced by the target culture, new tools for delivering lessons (interactive Java and Shockwave exercises) and testing, which are available to students any time they have the time or interest — 24 hours a day, 7 days a week. It is also an almost limitless publication outlet for my colleagues and I, not to mention my institution. Ultimately all course materials, including lecture notes, exercises, moot and credit testing, grading, and interactive exercises will be far more effective in conveying concepts that we have not even dreamed of yet.” After creating A Web of Online Dictionaries (WOD) in 1995, Robert Beard co-founded yourDictionary.com in 2000 as a web portal for online dictionaries and linguistic tools.

The Language Institute of the University of Hull, United Kingdom, created its Communications & Information Technology Centre (C&TI Centre) to provide information on how computer-assisted language learning could be integrated into existing courses. It also provided support to language teachers who were using computers in their teaching. According to June Thompson, manager of the C&TI Centre, interviewed in December 1998: “The internet has the potential to increase the use of foreign languages. The use of the internet has brought an enormous new dimension to our work of supporting language teachers in their use of technology in teaching. I suspect that for some time to come, the use of internet-related activities for languages will continue to develop alongside other technology-related activities (e.g. use of CD-ROMs — not all institutions have enough networked hardware). In the future I can envisage use of internet playing a much larger part, but only if such activities are pedagogy driven.”

Russon Wooldridge, professor at the Department of French Studies of the University of Toronto, Canada, wrote in February 2001: “My research, conducted once in an ivory tower, is now almost exclusively done through local or remote collaborations. All my teaching makes the most of internet resources (web and email): the two common places for a course are the classroom and the website of the course, where I put all course materials. I have published all my research data of the last twenty years on the web (re-edition of books, articles, texts of old dictionaries as interactive databases, 16th-century treaties, etc.). I publish proceedings of symposiums. I publish a journal. In May 2000, I organised an international symposium in Toronto on French studies enhanced by new technologies. I realise that without the internet I wouldn’t have as many activities, or at least they would be very different from the ones I have today. So I don’t see the future without them.”

The Massachusetts Institute of Technology (MIT) launched its OpenCourseWare (OCW) in September 2003 to put all its course materials online for free, after creating a pilot version one year earlier with 32 course materials. The MIT OpenCourseWare offered 500 course materials in March 2004, 1,400 course materials in May 2006, and all 1,800 course materials in November 2007. The course materials are regularly updated, and some of them are translated into Spanish, Portuguese and Chinese with the help of other organisations. An OpenCourseWare Consortium (later renamed

Open Education Consortium) was created in November 2005 as a global common project to offer the course materials of other educational institutions. One year later, it included the course materials of 100 universities worldwide.

---

## About resources for translators

The internet became a vital tool for translators, according to Marcel Grangier, head of the French Section of the Swiss government's Central Linguistic Services. He wrote in January 1999 in an email interview: "To work without the internet is simply impossible now. On top of all the tools we use (email, online press, services for translators), the internet is for us a vital and endless source of information in what I would call the 'non-structured sector' of the web. For example, when the answer to a translation issue cannot be found on websites presenting information in an organised way, in most cases search engines allow us to find the missing link somewhere on the network."

He explained in January 2000: "Our website was first conceived as an intranet service for translators in Switzerland, who often deal with the same kind of material as the translators of the federal government. Some parts of our website are useful to all translators, wherever they are. The section Dictionnaires Électroniques [Electronic Dictionaries] is only one section of the website. Other sections deal with public administration, legislation, the French language, and general information. Our website also hosts the pages of the Conference of Translation Services of European States (COTSOES)." Dictionnaires Électroniques offered links to monolingual dictionaries, bilingual dictionaries, multilingual dictionaries, abbreviations and acronyms, and geographical data, and was transferred to the new website of COTSOES in 2001.

Maria Victoria Marinetti, a Mexican engineer who was working as a Spanish-language teacher and translator in France, wrote in August 1999: "I have access to a large number of global information, which is very interesting for me. I can also regularly send or receive files back and forth. The internet allows me to receive or send general and technical translations from French into Spanish, and vice versa, and to correct texts in Spanish. In the technical or chemical fields, I offer technical assistance, as well as information on exporting high-tech equipment to Mexico or to other Latin American countries."

Praetorius, a language consultancy based in London, created Language Today, a magazine for linguists (translators, interpreters, terminologists, lexicographers, technical writers), with a print version and an online version. Geoffrey Kingscott, director of Praetorius, explained in September 1998: "We publish the print version of Language Today only in English, the common denominator of our readers. When we use an article which was originally in a language other than English, or report an interview which was conducted in a language other than English, we translate it into English and publish only the English version. This is because the number of pages we can print is constrained, governed by our customer base (advertisers and subscribers). But for our web edition we also give the original version."



Mailing lists were created to connect linguists across borders and languages, for example the Linguist List, created by Anthony Rodrigues Aristar and Helen Dry in 1990 at the University of Western Australia (60 subscribers) before moving to Texas A&M University in 1991. The list started its own website in 1997, and became a component of the WWW Virtual Library for linguistics.

Helen Dry, co-moderator of the list with Anthony Rodrigues Aristar, explained in August 1998: “The Linguist List has a policy of posting in any language, since it is a list for linguists. However, we discourage posting the same message in several languages, simply because of the burden extra messages put on our editorial staff. (We are not a bounce-back list, but a moderated one. So each message is organised into an issue with like messages by our student editors before it is posted.) Our experience has been that almost everyone chooses to post in English. But we do link to a translation facility that will present our pages in any of five languages. We also try to have at least one student editor who is genuinely multilingual, so that readers can correspond with us in languages other than English.”

She added in July 1999: “We are beginning to collect some primary data. For example, we have searchable databases of dissertation abstracts relevant to linguistics, of information on graduate and undergraduate linguistics programs, and of professional information about individual linguists. The dissertation abstracts collection is, to my knowledge, the only freely available electronic compilation in existence.”

---

## About terminology databases

NetGlos, which stands for Multilingual Glossary of Internet Terminology, was created in 1995 by the Worldwide Language Institute (WWLI) in Europe as an online voluntary collaborative project. NetGlos was available in 13 languages (Chinese, Croatian, Dutch/Flemish, English, French, German, Greek, Hebrew, Italian, Maori, Norwegian, Portuguese, Spanish) three years later.

Brian King, director of the Worldwide Language Institute, explained in September 1998 in an email interview: “Much of the technical terminology on the web is still not translated into other languages. And as we found with our Multilingual Glossary of Internet Terminology — known as NetGlos — the translation of these terms is not always a simple process. Before a new term becomes accepted as the ‘correct’ one, there is a period of instability where a number of competing candidates are used. Often an English loan word becomes the starting point — and in many cases the endpoint. But eventually a winner emerges that becomes codified into published technical dictionaries as well as the everyday interactions of the non-technical user. The latest version of NetGlos is the Russian one and it should be available in a couple of weeks or so. It will no doubt be an excellent example of the ongoing, dynamic process of ‘russification’ of web terminology.”

“Our NetGlos Project has depended on the goodwill of volunteer translators from Canada, the U.S., Austria, Norway, Belgium, Israel, Portugal, Russia, Greece, Brazil, New Zealand and other countries. I think the hundreds of visitors we get coming to the NetGlos pages every day is an excellent testimony to the success of these types of working relationships. I see the future

depending even more on cooperative relationships — although not necessarily on a volunteer basis. As a company that derives its very existence from the importance attached to languages, I believe the future will be an exciting and challenging one. But it will be impossible to be complacent about our successes and accomplishments. Technology is already changing at a frenetic pace. Lifelong learning is a strategy that we all must use if we are to stay ahead and be competitive. This is a difficult enough task in an English-speaking environment. If we add in the complexities of interacting in a multilingual/multicultural cyberspace, then the task becomes even more demanding. As well as competition, there is also the necessity for cooperation — perhaps more so than ever before. ”

Major international governmental organisations created free online versions of their terminology databases in 1997 and 1998, for example: (1) ILOTERM, a quadrilingual (English, French, German, Spanish) terminology database maintained by the International Labour Organisation (ILO), (2) TERMITE (Telecommunication Terminology Database), a quadrilingual (English, French, Spanish, Russian) terminology database maintained by the International Telecommunication Union (ITU), and (3) WHOTERM (WHO Terminology Information System), a trilingual (English, French, Spanish) terminology database maintained by the World Health Organisation (WHO).

Initially developed to assist in-house translators, Eurodicautom, the terminology database of the European Commission, launched in 1997 its free online version for the 15 member countries of the European Union and for linguists around the world. Eurodicautom was a multilingual database of economic, scientific, technical and legal terms and phrases, with language pairs for the 11 official languages of the European Union (Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish) and Latin. Eurodicautom had an average of 120,000 visitors per day in late 2003, when it announced a temporary hiatus (that lasted three years) to prepare for a larger terminology database in more languages with the enlargement of the European Union to 25 country members in May 2004 (and 27 country members in January 2007).

The project of a larger terminology database was studied as early as 1999 to merge the databases maintained by a number of European organisations (European Commission, European Parliament, Council of the European Union, Court of Justice, European Court of Auditors, European Economic and Social Committee, Committee of the Regions, European Investment Bank, European Central Bank, Translation Centre for the Bodies of the European Union).

The new terminology database, named IATE (InterActive Terminology for Europe), was available on the intranet of some European institutions in spring 2004. IATE was launched as a free public service on the internet in March 2007, with 1.4 million entries in the 23 official languages of the European Union (Bulgarian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, Swedish) and Latin. IATE offered 8.4 million words, 540,000 abbreviations and 130,000 phrases in 23 languages in 2010. IATE is maintained by the Translation Centre of the European Union Institutions in Luxembourg.

---

## About machine translation

According to the website of the European Association for Machine Translation (EAMT) in 1998: “Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another. One of the very earliest pursuits in computer science, MT has proved to be an elusive goal, but today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful for certain specific applications, usually in the domain of technical documentation. In addition, translation software packages which are designed primarily to assist the human translator in the production of translations are enjoying increasing popularity within professional translation organisations.”

As explained by Globalink, a company providing language translation software, on its website: “The computer uses three sets of data: the input text, the translation program and permanent knowledge sources (containing a dictionary of words and phrases of the source language), and information about the concepts evoked by the dictionary and rules for sentence development. These rules are in the form of linguistic rules for syntax and grammar, and some are algorithms governing verb conjugation, syntax adjustment, gender and number agreement and word re-ordering. Once the user has selected the text and set the machine translation process in motion, the program begins to match words of the input text with those stored in its dictionary. Once a match is found, the application brings up a complete record that includes information on possible meanings of the word and its contextual relationship to other words that occur in the same sentence. The time required for the translation depends on the length of the text. A three-page, 750-word document takes about three minutes to render a first draft translation.”

The website of Globalink also provided a history of machine translation. Here are a few excerpts: “From the very beginning, machine translation (MT) and natural language processing (NLP) have gone hand-in-hand with the evolution of modern computational technology. The development of the first general-purpose programmable computers during World War II was driven and accelerated by Allied cryptographic efforts to crack the German Enigma machine and other wartime codes. Following the war, the translation and analysis of natural language text provided a test bed for the newly emerging field of information theory. During the 1950s, research on automatic translation (known today as machine translation, or ‘MT’) took form in the sense of literal translation, more commonly known as word-for-word translations, without the use of any linguistic rules.”

“The Russian project initiated at Georgetown University in the early 1950s represented the first systematic attempt to create a demonstrable machine translation system. Throughout the decade and into the 1960s, a number of similar university and government-funded research efforts took place in the United States and Europe. At the same time, rapid developments in the field of theoretical linguistics, culminating in the publication of Noam Chomsky’s ‘Aspects of the Theory of Syntax’ (1965), revolutionised the framework for the discussion and understanding of the phonology, morphology, syntax and semantics of human language. In 1966, the U.S. government-issued ALPAC [Automatic Language Processing Advisory Committee] report offered a prematurely negative assessment of the value and prospects of practical machine translation systems, effectively putting an end to funding and experimentation in the field for the next decade.”

“It was not until the late 1970s, with the growth of computing and language technology, that serious efforts began once again. This period of renewed interest also saw the development of the transfer model of machine translation and the emergence of the first commercial MT systems.

While commercial ventures such as Systran and Metal began to demonstrate the viability, utility and demand for machine translation, these mainframe-bound systems also illustrated many of the problems in bringing MT products and services to market. High development cost, labour-intensive lexicography and linguistic implementation, slow progress in developing new language pairs, inaccessibility to the average user, and inability to scale easily to new platforms are all characteristics of these second-generation systems.”

In “Web embraces language translation”, an article published by ZDNN (ZD Network News) on 21 July 1998, journalist Martha L. Stone explained: “While these so-called ‘machine’ translations are gaining worldwide popularity, company execs admit they’re not for every situation. Representatives from Globalink, Alis and Systran use such phrases as ‘not perfect’ and ‘approximate’ when describing the quality of translations, with the caveat that sentences submitted for translation should be simple, grammatically accurate and idiom-free. ‘The progress on machine translation is moving at Moore’s Law — every 18 months it’s twice as good,’ said Vin Crosbie, a web industry analyst in Greenwich, Conn. ‘It’s not perfect, but some people don’t realise I’m using translation software.’ With these translations, syntax and word usage suffer, because dictionary-driven databases can’t decipher between homonyms — for example, ‘light’ (as in the sun or light bulb) and ‘light’ (the opposite of heavy). Still, human translation would cost between [US]\$50 and \$60 per web page, or about 20 cents per word, Systran’s Sabatakakis said. While this may be appropriate for static ‘corporate information’ pages, the machine translations are free on the web, and often less than \$100 for software, depending on the number of translated languages and special features.”

IBM released a high-end professional product, the WebSphere Translation Server, in March 2001. The software could instantly translate web pages, emails and chats from/into eight languages (Chinese, English, French, German, Italian, Japanese, Korean, Spanish), with 500 words processed per second. Machine translation software was also created by Systran, Alis Technologies, Lernout & Hauspie (who bought Globalink) and Softissimo.

Another experiment was led by the Pan American Health Organisation (PAHO), based in Washington D.C. The PAHO Translation Unit was one of the first ones to use two in-house machine translation systems — SPANAM (Spanish to English) developed since 1980, and ENGSPAN (English to Spanish) developed since 1985. In-house and free-lance translators could post-edit the raw output to produce quality translations with a 30-50 percent gain in productivity. SPANAM and ENGSPAN were available on PAHO’s LAN (Local Area Network) for the technical and administrative staff, and in PAHO field offices. SPANAM and ENGSPAN were also licensed to public and non-profit institutions in the United States, Latin America and Spain. The software was later renamed PAHOMTS, and new versions were added, with English-Portuguese (July 2003), Portuguese-English (July 2003), Spanish-Portuguese (March 2004) and Portuguese-Spanish (June 2005).

---

## About computer-assisted translation

A different project was led by the Computer-assisted Translation and Terminology (CTT) Unit of the World Health Organisation (WHO). While maintaining its trilingual (English, French, Spanish) terminology database WHOTERM (WHO Terminology Information System), the CTT Unit was assessing technical options for using computer-assisted translation systems based on translation memory.

While machine translation is the automated process of translating from one language to another language, with no human intervention during the translation process, computer-assisted translation (CAT) requests interactions between the translator and the software during the translation process. CAT software is based on translation memory, with terminology processing in real time, i.e. the instant access to previous translations for portions of text, that can be accepted, rejected or modified, and the new translation added to the memory.

Wordfast, a popular CAT tool created in 1999 by Yves Champollion, was compatible with other main MT or CAT software such as the IBM Websphere Translation Server and SDL Trados. Available for any platform (Windows, Mac, Linux, etc.), Wordfast had 14,000 customers worldwide in 2010, including the United Nations, NASA (National Aeronautics and Space Administration), Sony, Coca-Cola and McGraw-Hill.

According to Tim McKenna, a mathematics teacher and writer interviewed in October 2000: “When software gets good enough for people to chat or talk on the web in real time in different languages, then we will see a whole new world appear before us. Scientists, political activists, businesses and many more groups will be able to communicate immediately without having to go through mediators or translators.”

Randy Hobler, a marketing consultant for translation software and services, described the next steps in September 1998 in an email interview: “We are rapidly reaching the point where highly accurate machine translation of text and speech will be so common as to be embedded in computer platforms, and even in chips in various ways. At that point, and as the growth of the web slows, the accuracy of language translation hits 98%, and the saturation of language pairs has covered the vast majority of the market, language transparency (any-language-to-any-language communication) will be too limiting a vision for those selling this technology.”

“The next development will be ‘transcultural, transnational transparency’, in which other aspects of human communication, commerce and transactions beyond language alone will come into play. For example, gesture has meaning, facial movement has meaning and this varies among societies. The thumb-index finger circle means ‘OK’ in the United States. In Argentina, it is an obscene gesture. When the inevitable growth of multimedia, multilingual videoconferencing comes about, it will be necessary to ‘visually edit’ gestures on the fly. The MIT Media Lab, Microsoft and many others are working on computer recognition of facial expressions, biometric access identification via the face, etc. It won’t be any good for a U.S. business person to be making a great point in a web-based multilingual video conference to an Argentinian, having his words translated into perfect Argentinian Spanish if he makes the ‘O’ gesture at the same time. Computers can intercept this kind of thing and edit them on the fly.”

“There are thousands of ways in which cultures and countries differ, and most of these are computerisable to change as one goes from one culture to the other. They include laws, customs, business practices, ethics, currency conversions, clothing size differences, metric versus English

system differences, etc. Enterprising companies will be capturing and programming these differences and selling products and services to help the peoples of the world communicate better. Once this kind of thing is widespread, it will truly contribute to international understanding.”

---

## About free machine translation services

The search engine AltaVista launched AltaVista Translation — better known as Babel Fish — in December 1997 as the first free machine translation service from English to five other languages (French, German, Italian, Portuguese, Spanish), and vice versa. The original web page and the instant translation were face-to-face on the screen. Translating any short text was also possible by copying and pasting it into the interface. The result was far from perfect, but helpful and free, and Babel Fish contributed to a plurilingual web.

Babel Fish was developed by Systran (System Translation), a company specialising in automated language solutions. As explained on Systran’s website: “Machine translation (MT) software translates one natural language into another natural language. MT takes into account the grammatical structure of each language and uses rules to transfer the grammatical structure of the source language (text to be translated) into the target language (translated text). MT cannot replace a human translator, nor is it intended to.”

In “Web embraces language translation”, an article published by ZDNN (ZD Network News) on 21 July 1998, journalist Martha L. Stone explained: “Systran has partnered with AltaVista and reports between 500,000 and 600,000 visitors a day on [babelfish.altavista.digital.com](http://babelfish.altavista.digital.com), and about 1 million translations per day — ranging from recipes to complete web pages. About 15,000 sites link to Babel Fish, which can translate to and from French, Italian, German, Spanish and Portuguese. The site plans to add Japanese soon. ‘The popularity is simple. With the internet, now there is a way to use U.S. content. All of these contribute to this increasing demand,’ said Dimitros Sabatakakis, group CEO of Systran, speaking from his Paris home.” Babel Fish moved on Yahoo!’s website in May 2008, and was replaced by Microsoft’s Bing Translator in May 2012.

Google Translate is a free online language translation service launched in October 2007, that instantly translates a text or web page into another language. Users paste a text in the web interface or supply a hyperlink. The machine translation is produced by statistical analysis instead of the traditional rule-based analysis. As stated on its website: “Google Translate can help users understand the general content of a foreign language text, but doesn’t deliver accurate translations.”

Prior to this date, Google used a Systran-based translation service, with several stages in languages pairs detailed on Wikipedia: (1) English to French, German, and Spanish, and vice versa; (2) English to Portuguese and Dutch, and vice versa; (3) English to Italian, and vice versa; (4) English to simplified Chinese, Japanese and Korean, and vice versa; (5) English to Arabic, and vice versa (in April 2006); (6) English to Russian, and vice versa (in December 2006); (7) English to traditional Chinese, and simplified Chinese to traditional Chinese, and vice versa (in February 2007).

Google Translate's developments in 2007-10 were: (8) all language pairs previously available, in any language combination (in October 2007); (9) English to Hindi, and vice versa (in December 2007); (10) Bulgarian, Croatian, Czech, Danish, Finnish, Greek, Norwegian, Polish, Romanian, Swedish, with any combination (in May 2008); (11) Catalan, Filipino, Hebrew, Indonesian, Latvian, Lithuanian, Serbian, Slovak, Slovene, Ukrainian, Vietnamese (in September 2008); (12) Albanian, Estonian, Galician, Hungarian, Maltese, Thai, Turkish (in January 2009); (13) Persian (in June 2009); (14) Afrikaans, Belarusian, Icelandic, Irish, Macedonian, Malay, Swahili, Welsh, Yiddish (in August 2009); (15) Haitian Creole (in January 2010); (16) Armenian, Azeri, Basque, Georgian, Urdu (in May 2010); (17) Latin (in October 2010).

A speech program to read the translated text was available in 2009. Different translations for the same word were available in 2011. The Google Translator Toolkit was available in June 2009 as a free web service for (human) translators to edit the translations generated by Google Translate, with English as a source language and 47 target languages. Translators could also share translations, and create glossaries and translation memories.

---

## About the Ethnologue

The "Ethnologue: Languages of the World", a reference catalogue published in print by SIL International, launched its free online version in 1996, with a full description of the 6,700 living languages spoken in 228 countries.

As explained in January 2000 by Barbara Grimes, editor of the Ethnologue since 1971, in an email interview: "The Ethnologue is a catalogue of the languages of the world, with information about where they are spoken, an estimate of the number of speakers, what language family they are in, alternate names, names of dialects, other socio-linguistic and demographic information, dates of published Bibles, a name index [the Ethnologue Name Index], a language family index [the Ethnologue Language Index], and language maps. (...) We have had requests for the Ethnologue in a few other languages, but we do not have the personnel or funds to do the translation or maintenance, since it is constantly being updated."

The Ethnologue was founded in 1951 by Richard Pittman as a catalogue of minority languages, to share information on language development needs with his colleagues at SIL International (formerly known as the Summer Institute of Linguistics) and with other language researchers around the globe. Information was expanded from minority languages to include all known languages of the world in 1971, with the help of thousands of linguists from partner organisations. Barbara Grimes supervised in 1967-73 an in-depth revision of the information available for Africa, the Americas, the Pacific and part of Asia. The number of identified languages grew from 4,493 to 6,809, with more information recorded on each language in the computer database created in 1971.

What exactly is a language? According to the website of the Ethnologue: "How one chooses to define a language depends on the purposes one has in identifying one language as being distinct from another. Some base their definition on purely linguistic grounds, focusing on lexical and grammatical differences. Others may see social, cultural, or political factors as being primary. In

addition, speakers themselves often have their own perspectives on what makes a particular language uniquely theirs. Those are frequently related to issues of heritage and identity much more than to the actual linguistic features. In addition, it is important to recognise that not all languages are oral. Sign languages constitute an important class of linguistic varieties that merit consideration.”

A new edition of the Ethnologue was published approximately every four years until 2012. Since 2013, the Ethnologue has been published every year, with the online version available before the printed version, to keep up with the fast pace of the internet. The Ethnologue identified 6,909 living languages in 2009, 7,102 living languages in 2015, and 7,099 living languages in 2017. A paid subscription model for extensive users in high income countries was created in 2016 in order to sustain the Ethnologue project.

At the invitation of the International Organisation for Standardisation (ISO) in 2002, SIL International worked on the new standard ISO 639-3 (2007), that reconciled the complete set of three-letter identifiers used in the Ethnologue since the inception of its database in 1971 with the 400 three-letter codes used in the previous standard ISO 639-2 (1998), as well as other three-letter codes developed by the Linguist List for ancient and constructed languages. (The first ISO standard to identify languages was ISO 639-1 (1988) as a set of two-letter language codes.)

Approved in 2006 and published in 2007, ISO 639-3 (2007) has provided three-letter codes for identifying 7,589 languages (living and extinct, ancient and reconstructed, major and minor, written and unwritten), including sign languages. SIL International was named the registration authority for the language identifier inventory, and has administered the annual cycle of changes and updates since then.

---

## About minority languages

Guy Antoine, a Haitian-American software developer born in Haiti and living in New York, founded the website Windows on Haiti in April 1998 on his free time to promote Haitian Creole, a French-based creole language spoken not only in Haiti but also in the Dominican Republic, the United States, Canada, and other countries.

Guy Antoine wrote in June 2001 in an email interview: “Who are the Haitian people without Kreyól [Haitian Creole], the language that has evolved and bound various African tribes transplanted in Haiti during the slavery period? It is the most palpable exponent of commonality that defines us as a people. However, it is primarily a spoken language, not a widely written one. I see the web changing this situation more than any traditional means of language dissemination. Our site aims to be a major source of information about Haitian culture, and a tool to counter the persistently negative images of Haiti from the traditional media. The scope of this effort extends beyond mere commentary to the diversity of arts and history, cuisine and music, literature and reminiscences of traditional Haitian life. In short, the site opens some new windows to the culture of Haiti.”



“The primary language of Windows on Haiti is English, but one will equally find a center of lively discussion conducted in Kreyól. In addition, one will find documents related to Haiti in French, in the old colonial Creole, and I am open to publishing others in Spanish and other languages. I do not offer any sort of translation, but multilingualism is alive and well at the site, and I predict that this will increasingly become the norm throughout the web. Kreyól is the only national language of Haiti, and one of its two official languages, the other being French. It is hardly a minority language in the Caribbean context, since it is spoken by eight to ten million people. I have created two discussion forums on my website, held exclusively in Kreyól. One is for general discussions just about everything but obviously more focused on Haiti’s current sociopolitical problems. The other is reserved only to debates of writing standards for Kreyól. Those debates have been quite spirited with the participation of a number of linguistic experts. The uniqueness of these forums is their non-academic nature.”

According to some linguists, a language dies every 14 days. A good way to counter it is to bring together language communities via the internet, to help revitalise these languages through digital technology, and to strengthen the presence of language communities in social media.

Kevin Scannell, a computer scientist and professor at Saint Louis University, Missouri, created the website Indigenous Tweets on his free time to identify tweets in indigenous and minority languages. He designed An Crúbadán, a statistical software crawling the web to find Twitter threads. Indigenous Tweets identified 35 languages in March 2011, 71 languages in April 2011, 144 languages in March 2013, and 184 languages in October 2017.

Indigenous Tweets’ home page lists all languages identified as being active on Twitter. People click on the corresponding row of the language they are interested in, and are redirected to a new page that lists users in that language (with a maximum of 500 users) and statistics for each user: number of tweets, number of followers, percentage of tweets in the given language (some users tweet in both a global language and a minority language), and date of the latest tweet. The main minority languages are Haitian Creole (with users from the Caribbean, North America and other places), Basque and Welsh. People can also get in touch directly via Twitter. A number of joint projects started this way.

As explained by Kevin Scannell on his blog: “Together we’re breaking down the idea that only global languages like English and French have a place online! The primary aim of Indigenous Tweets is to help build online language communities through Twitter. We hope that the site makes it easier for speakers of indigenous and minority languages to find each other in the vast sea of English, French, Spanish, and other global languages that dominate Twitter. Even speakers of languages like Basque and Welsh with vibrant online communities have been surprised to find just how many people there are tweeting in their language. This is the other goal of Indigenous Tweets: it’s a message to the world that says ‘We are here and we’re proud of our languages’. For languages with just a few users, I hope it inspires some people to start — make your voice heard!”

Kevin Scannell created a second website, Indigenous Blogs, in September 2011 to identify blogs written in indigenous and minority languages, and to offer a similar platform for people to get in touch. He began with blogs hosted by Blogspot (which also hosts his own blog), WordPress and Tumblr. Indigenous Blogs identified blogs in 50 languages in September 2011, in 74 languages in March 2013, and in 85 languages in October 2017.

---

# About endangered languages

UNESCO (United Nations Educational, Scientific and Cultural Organisation) published in 2010 the free online version of its “Atlas of the World’s Languages in Danger” alongside its printed edition (3rd edition, 2010), edited by Christopher Moseley. The previous editions of the atlas in 1996 and 2001 only existed in print.

Available in three languages (English, French, Spanish) like the printed edition, the online atlas included 2,473 languages in 2010 and 2,464 languages in 2017. It can be searched by country and area, language name, number of speakers from/to, language vitality, and ISO 639-3 code. The alternate names of the languages (spelling variants, dialects, names in non-Roman scripts) are also provided.

UNESCO experts have established six degrees (safe, vulnerable, definitely endangered, severely endangered, critically endangered, extinct) to define the vitality or endangerment of a language. (1) “Safe” — not included in the atlas — means that the language is spoken by all generations and that inter-generational transmission is uninterrupted. (2) “Vulnerable” means that most children speak the language, but it may be restricted to certain places, for example at home. (3) “Definitely endangered” means that children no longer learn the language as a native language at home. (4) “Severely endangered” means that the language is spoken by grandparents and older generations; the parent generation may understand it, but doesn’t use it with their children or among themselves. (5) “Critically endangered” means that the youngest speakers are grandparents and older, who speak the language partially and infrequently. (6) “Extinct” means that there are no speakers left; the atlas includes languages that are presumably extinct since the 1950s.

When exactly is a language considered endangered? As explained on the website of the atlas: “A language is endangered when its speakers cease to use it, use it in fewer and fewer domains, use fewer of its registers and speaking styles, and/or stop passing it on to the next generation. No single factor determines whether a language is endangered.” UNESCO experts have identified nine factors to be considered: (1) inter-generational language transmission; (2) absolute number of speakers; (3) proportion of speakers within the total population; (4) shifts in domains of language use; (5) response to new domains and media; (6) availability of materials for language education and literacy; (7) governmental and institutional language attitudes and policies including official status and use; (8) attitudes of community members towards their own language; (9) amount and quality of documentation.

When and why do languages disappear? “A language disappears when its speakers disappear or when they shift to speaking another language — most often, a larger language used by a more powerful group. Languages are threatened by external forces such as military, economic, religious, cultural or educational subjugation, or by internal forces such as a community’s negative attitude towards its own language. Today, increased migration and rapid urbanisation often bring along the loss of traditional ways of life and a strong pressure to speak a dominant language that is — or is perceived to be — necessary for full civic participation and economic advancement.”

The UNESCO atlas classifies Gaelic as a “definitely endangered” language. There were 59,000 Gaelic speakers (over 1 percent of the population), according to the 2011 census. These figures were much lower than the 200,000 Gaelic speakers (4.5 percent of the population) in the 1901 census.

This has not always been the case. For many centuries, everyone spoke Gaelic in Scotland and Ireland, and scholars disseminated their writings in Gaelic throughout Europe. Over the centuries, English gradually became the dominant language, including on the Scottish Western Isles, despite the presence of Scottish Gaelic as the first community language. The revival of Gaelic culture dates back to the early 19th century, in the form of poetry, prose and music. Between the two world wars, a radio channel began broadcasting the news in Gaelic, and Gaelic began being learned again in schools. Today, more novels are published in Gaelic than at any other time. Radio nan Gàidheal has broadcast in Gaelic since the 1980s, and the TV channel ALBA has offered shows in Gaelic since the early 2000s. Both have a web presence, which has boosted their audience. Wikipedia has its Gaelic version, named Uicipeid.

Michael Bauer, a freelance translator from English to Scottish Gaelic, has worked on several localisation projects on his free time, “just for the love of it”, with a fellow localiser who on the web only goes by GunChleoc (“a woman” in Scottish Gaelic), a proof that few people can do a lot for their language community. The localisation projects included the Gaelic versions of the web browser Opera (as early as 2001), Firefox (Mozilla web browser), Thunderbird (Mozilla messaging), Lightning (Mozilla calendar), Google Chrome, OpenOffice, LibreOffice, the VLC media player, the game Freeciv (the open source version of the game Civilisation), and Accentuate.us (a software inserting accents). Michael Bauer also created the spell checker An Dearbhair Beag with Kevin Scannell. Since 2012, he has worked on a few paid projects with GunChleoc, for example the Gaelic language packs for Microsoft Windows and Microsoft Office.

There are three online dictionaries in Scottish Gaelic. The first dictionary is Stòr-dàta, an online dictionary which is mostly a word list managed by Sabhal Mòr Ostaig, a college on the Isle of Skye where all the courses are taught in Scottish Gaelic. The second dictionary is the Dwelly, a Gaelic dictionary published in 1911, which is to Gaelic what the Oxford English Dictionary is to English. Its digital edition is the result of a ten-year labour of love by Michael Bauer and his colleague Will Robertson. The third dictionary is Am Faclair Beag, which means “small dictionary” but is actually a large dictionary offering both the Dwelly and more modern data, also created and maintained by Michael Bauer and Will Robertson.

Michael Bauer wrote in October 2015 in an email interview: “There are, sadly, far too few users and there are some aspects which actually actively limit usage. For example, Gaelic schools cannot install Gaelic software because the IT contracts are given by the councils to outside IT companies who only provide English software and operating systems. Because they limit the admin rights of the users at schools, this means it is very difficult to install software which is not on their official ‘list’ and because Gaelic is not mentioned in the contract, they don’t put it there. Free and open software has helped carve out more of a space on the web for Gaelic, and cooperating with commercial long-term partners is helping to produce some very useful enabling technologies such as the predictive texting tool Adaptxt or the upcoming text-to-speech tool with Cereproc.”

“A central storage space for translations would be useful for localisation projects, with a shared translation memory, thus avoiding to endlessly retranslate the same terms, phrases and sentence segments. If the translations could be available from the same site, like a meta-Pootle [a community localisation server], everyone working for the revival of a minority language on the web would benefit from it. There actually was/is something a bit like that, Ubuntu’s Launchpad, but unfortunately there is not enough coordination between Launchpad and the projects and much effort is going to waste by people working on Launchpad and the translations not going anywhere. There is also AmaGama these days which is something like that but not commonly used apart from some like Mozilla and LibreOffice (I think). Part of the problem is there are so many platforms these days, all trying to carve out a niche... some of them commercial, like Transifex or Crowdin.”

What is the best way to help language revitalisation efforts? Many minority, indigenous and endangered languages still need language dictionaries, grammars and glossaries. Some of them even need basic language technologies such as keyboard settings or spell checkers.

As explained on his blog by Kevin Scannell, founder of the websites Indigenous Tweets and Indigenous Blogs in 2011: “Speakers of indigenous and minority languages around the world are struggling to keep their languages and cultures alive. More and more language groups are turning to the web as a tool for language revitalisation, and as a result there are now thousands of people blogging and using social media sites like Facebook and Twitter in their native language. These sites have allowed sometimes scattered communities to connect and use their languages online in a natural way. Social media have also been important in engaging young people, who are the most important demographic in language revitalisation efforts.”

---

## Timeline

[year-month]

1963: ASCII (American Standard Code for Information Interchange) is the first encoding system for English (and Latin).

1971-07: Michael Hart sends a link to eText #1 to the 100 users of the pre-internet. Project Gutenberg is born.

1974: Vinton Cerf and Robert Kahn invent the communication protocols at the heart of the internet.

1977: The International Federation of Library Associations (IFLA) creates UNIMARC as a common bibliographic format for library catalogues.

1983: After being a network linking U.S. governmental agencies, universities and research centres, the internet starts its progression worldwide.

1984: The copyleft, invented by Richard Stallman, ensures that computer software can be freely used and improved by using a GPL (General Public License).

1990: Tim Berners-Lee invents the World Wide Web and gives his invention to the world.

1990: Anthony Rodrigues Aristar and Helen Dry create the Linguist List.

1991-01: The Unicode Consortium is founded to develop Unicode, a universal encoding system for the processing, storage and interchange of text data in any language.

1992: The Internet Society (ISOC) is founded by Vinton Cerf to promote the development of the

internet.

1992: Paul Southworth creates the Etext Archives as a home for electronic texts of any kind.

1992: Projekt Runeberg is the first Swedish online collection of public domain books.

1993: John Mark Ockerbloom creates The Online Books Page to facilitate access to books that are available for free on the internet.

1993-04: ABU-La Bibliothèque Universelle (ABU-The Universal Library) is the first French online collection of public domain books.

1993-06: Adobe launches PDF (Portable Document Format), Acrobat Reader to read PDF files, and Adobe Acrobat to make them.

1993-07: John Labovitz creates The E-Zine-List as a list of electronic zines around the world.

1993-11: Mosaic is the first web browser for the general public.

1994: Netscape Navigator is the second web browser for the general public.

1994: Projekt Gutenberg-DE is the first German online collection of public domain books.

1994: Michel Martin creates Travlang as a list of free online translation dictionaries for travellers.

1994: Pierre Perroud creates Athena, the first Swiss online collection of public domain books.

1994-05: Tyler Chambers creates The Human-Languages Page (H-LP) as a catalogue of online linguistic resources.

1994-07: Internet users with a native language other than English reach 5 percent.

1994-10: The World Wide Consortium (W3C) is founded to develop protocols for the web.

1995: The Worldwide Language Institute (WWLI) launches NetGlos, a multilingual online glossary of internet terminology.

1995: Microsoft launches its web browser Internet Explorer.

1995: Robert Beard creates A Web of Online Dictionaries as a directory of free online dictionaries.

1995: Tyler Chambers creates the Internet Dictionary Project (IDP) as a collaborative project to create free online bilingual dictionaries.

1995-07: Jeff Bezos launches the online bookstore Amazon.com.

1996: The Ethnologue, a reference catalogue for all living languages, creates its free online version.

1996-04: The Internet Archive is founded by Brewster Kahle to archive the web and other digital content for present and future generations.

1996-04: Robert Ware creates OneLook Dictionaries as a fast finder in hundreds of free online dictionaries.

1996-12: Member states of the World Intellectual Property Organisation (WIPO) adopt the WIPO Copyright Treaty (WCT).

1997: The Dictionnaire Universel Francophone (French-language Universal Dictionary) from Hachette is freely available online.

1997: The French National Library launches its digital library Gallica.

1997-01: The International Labour Office (ILO) organises its first symposium on multimedia convergence.

1997-01: Several European national libraries create a common website named Gabriel.

1997-04: There are one million websites worldwide.

1997-05: The British Library launches its first OPAC (Online Public Access Catalogue).

1997-08: O'Reilly Japan publishes "For a Multilingual Web" by Yoshi Mikami in Japanese, and translates it into English, French and German the following year.

1997-09: The Internet Bookshop (United Kingdom) starts selling books published in the United States.

1997-12: The Italian translation company Logos puts all its linguistic resources (dictionaries,

glossaries, grammars, conjugators) online for free.

1997-12: Yahoo!, a directory of websites, offers its home page in seven languages for a growing number of non-English-speaking users.

1997-12: The search engine AltaVista launches Babel Fish, its free instant online translation service.

1997-12: There are 70 million internet users — 1.7 percent of the world population.

1998-07: Internet users with a native language other than English reach 25 percent.

1998-10: Amazon creates its first two subsidiaries in the United Kingdom and Germany.

1998-10: The Digital Millennium Copyright Act (DMCA) extends copyright to 70 years after the author's death.

1999: Michael Kellogg creates WordReference.com to offer free online bilingual dictionaries and discussion forums for linguists.

1999-09: The Open eBook (OeB) is published as a standard format for e-books.

1999-12: Britannica.com is the online version of the Encyclopaedia Britannica, available for free and then for a fee.

1999-12: The French-language Encyclopaedia Universalis creates its online version, with a paid subscription and some free articles.

2000-01: The wiki becomes popular as a collaborative website.

2000-01: The Million Book Project wants to offer one million free e-books in several languages, later included in the Internet Archive.

2000-02: Robert Beard co-founds yourDictionary.com as a web portal for dictionaries and other linguistic resources.

2000-03: The Oxford English Dictionary (OED) is available online, with a paid subscription.

2000-03: There are 300 million internet users — 5 percent of the world population.

2000-07: Internet users with a native language other than English reach 50 percent.

2000-08: Amazon launches its French subsidiary Amazon.fr.

2000-09: Quebec's GDT (Grand Dictionnaire Terminologique – Large Terminology Dictionary) is a major bilingual French-English online dictionary available for free.

2001: Lawrence Lessig creates Creative Commons for authors to be able to share their work on the internet, and for creators to be able to copy and remix it.

2001-01: Jimmy Wales and Larry Sanger create Wikipedia as a free collaborative online encyclopedia.

2001-03: IBM launches the WebSphere Translation Server to handle machine translation on a large scale in eight languages.

2001-04: There are 17 million PDAs and 100,000 e-readers worldwide, according to a Seybold Report.

2001-05: The European Union Copyright Directive (EUCD) extends copyright from 50 to 70 years after the author's death.

2001-10: The Internet Archive launches the Wayback Machine to crawl the 30 billion web pages archived since 1996.

2002-02: The Budapest Open Access Initiative (BOAI) is signed as the founding text of the open access movement for free access to research literature.

2002-03: The Oxford Reference Online (ORO) is launched as a major online encyclopedia conceived for the web, with a paid subscription.

2002-12: Creative Commons publishes its first licenses.

2003-09: The Massachusetts Institute of Technology (MIT) creates its OpenCourseWare (OCW) to offer all its course materials for free on the web.

2003-10: The Public Library of Science (PLOS) launches its first scientific and medical journals, with all its articles under a Creative Commons license.

2003-12: One million works on the internet use a Creative Commons license.

2004: Tim O'Reilly, founder of O'Reilly Media, coins the term "web 2.0" as the title of a conference he is organising.

2004-01: The European Library replaces Gabriel as the web portal for European national libraries.

2004-02: Mark Zuckerberg creates Facebook for his fellow students before extending it to the world.

2004-03: The Research Libraries Group (RLG) launches RedLightGreen as the first free online multilingual union catalogue for libraries.

2004-05: The European Union now has 20 official languages (instead of 11 languages and Latin) after its enlargement.

2004-10: Google launches Google Print (the future Google Books).

2005-04: The International Digital Publishing Forum (IDPF) replaces the Open eBook Forum.

2005-10: The Internet Archive launches the Open Content Alliance (OCA) to offer a worldwide public digital library in many languages.

2005-12: The Massachusetts Institute of Technology (MIT) launches its OpenCourseWare Consortium for other universities to offer their course materials online for free.

2005-12: There are one billion internet users — 15.7 percent of the world population.

2006: There are 90 million smartphones and one billion mobile phones worldwide.

2006-06: Twitter is launched as a micro-blogging tool with 140-character messages.

2006-08: Google replaces Google Print with Google Books.

2006-08: OCLC's union catalogue WorldCat launches its free online version.

2006-10: Microsoft creates Live Search Books before giving its collection of e-books to the Internet Archive two years later.

2006-11: There are 100 million websites.

2006-12: Gallica, the digital library of the French National Library, offers 90,000 books and 80,000 images.

2007-01: The European Union has 23 official languages instead of 20, with Bulgarian, Irish and Romanian.

2007-02: Creative Commons publishes the versions 3.0 of its licenses, with an international license and compatibility with other licenses like copyleft and GPL.

2007-03: Larry Sanger creates Citizendium as a free collaborative online encyclopedia led by experts.

2007-04: yourDictionary.com offers a directory of 2,500 dictionaries and grammars in 300 languages.

2007-06: The European Commission launches the free public version of its multilingual terminology database IATE (InterActive Terminology for Europe).

2007-09: The International Digital Publishing Forum (IDPF) publishes the first version of EPUB to replace the OeB (Open eBook) format.

2007-10: Google launches Google Translate as its free online translation service, after using Systran's translation service for two years.

2007-12: Unicode (created in 1991) supersedes ASCII (created in 1963) as the main encoding system on the internet.

2008-07: PDF is released as an open standard and published by the International Organisation for Standardisation (ISO) as ISO 32000-1:2008.

2008-11: Europeana is launched as the European public digital library.

2009-06: The Google Translator Toolkit is a free web service for (human) translators to edit machine translations produced by Google Translate.  
2010-06: Facebook celebrates its 500 million users.  
2010-12: 400 million works on the internet use a Creative Commons license.  
2011-01: Wikipedia celebrates its tenth anniversary with 17 million articles in 270 languages.  
2011-03: There are 2 billion internet users — 30.2 percent of the world population.  
2011-03: Kevin Scannell creates the website Indigenous Tweets to identify tweets in minority and indigenous languages.  
2013: The Ethnologue, a reference catalogue for all living languages, publishes its free online version before its paid print version.  
2013-11: Creative Commons publishes the versions 4.0 of its licenses.  
2014-12: 882 million works on the internet use a Creative Commons license.  
2015-01: There are 7,102 living languages, according to the Ethnologue.  
2015-03: There are 3 billion internet users — 42.3 percent of the world population.  
2015-04: The Online Books Page gives access to two million free e-books.  
2015-05: There are one billion websites.  
2015-07: Internet users with a native language other than English reach 75 percent.

---

Copyright © 2015-19 [Marie Lebert](#)  
[License CC BY-NC-SA version 4.0](#)

Written by marielebert

2015-11-18 at 21:49

Posted in [Uncategorized](#)

**Blog at WordPress.com.**



# Languages on the web – a timeline



By Marie Lebert, version of 3 May 2025.

After the invention of the web in 1990, internet users with a native language other than English reached 5 percent in 1994, 20 percent in 1998, 50 percent in 2000, and 75 percent in 2015. Hundreds of people helped promote both their own and other languages and cultures — sometimes in their spare time, and often using English as a lingua franca — for the web to become truly multilingual. Translators, librarians, language teachers and various others teamed up with web and software developers to create localised websites, online dictionaries, terminology databases, language reference tools, and other bilingual or plurilingual content. I interviewed some of them (same edition here) over the years, by email and in person, for them to describe how they contributed to a multilingual web. While their names are mostly unknown, their work was groundbreaking (with more interviews included in these ebooks) and paved the way for the web we use every day.

---

\* [French version]

\* [Spanish version]

---

## **[pre-web] 1963:**

**ASCII (American Standard Code for Information Interchange) was the first text encoding standard – based on the English alphabet. [#1]**

Published by the American Standards Association (ASA) in 1963, with a major revision in 1967, ASCII (American Standard Code for Information Interchange) is a 7-bit character encoding standard with 128 code points and 95 printable characters (digits 0 to 9, lowercase letters a to z, uppercase letters A to Z, punctuation symbols) based on the English alphabet for information interchange in English (and Latin).

With computer technology spreading outside North America, 8-bit variants of ASCII known as extended ASCII were published in 1987 with up to 128 additional characters in order to process the accented characters of other languages using the Latin alphabet, for example French, Spanish or German.

---

## **[pre-web] July 1971:**

**Michael Hart typed a short text that became eText #1 of Project Gutenberg, the first digital library offering free digitised editions of literary classics in the public domain. [#2]**

Founded in July 1971 by Michael Hart, a student at the University of Illinois, when the internet was just a glimmer, Project Gutenberg was a visionary project to create electronic versions of literary classics in the public domain and to disseminate them for free. In the 16th century, Johannes Gutenberg allowed anyone to have a few print books for a small cost (compared to the cost of manuscripts). In the 20th century, Project Gutenberg would allow anyone to have a full digital library at no cost.

After typing a few short texts alone, Michael Hart was joined by a small team of volunteers who patiently typed some literary classics page after page. The invention of the web in 1990 made it easier to distribute ebooks and to recruit new volunteers. The first digitised editions of translated literary classics (English translations of French works) were added in 1993, followed by the first literary classics in non-English languages in 1997.

“We consider e-text to be a new medium, with no real relationship to paper, other than presenting the same material, but I don’t see how paper can possibly compete once people each find their own comfortable way to e-texts, especially in schools.” — Michael Hart, founder of Project Gutenberg, interviewed in August 1998.

---

#### **[pre-web] 1974:**

**Vinton Cerf and Robert Kahn invented the communication protocols at the heart of the internet. [#3]**

The internet took off in 1974 after the creation of the TCP/IP protocols (TCP/IP: Transmission Control Protocol / Internet Protocol) by computer scientists Vinton Cerf and Robert Kahn. It expanded as a network linking governmental agencies, universities and research centres in North America. It got its first boost with the invention of the web by computer scientist Tim Berners-Lee in 1990, and its second boost with the distribution of Mosaic in 1993 as the first web browser for the general public.

Often named the Father of the internet, Vinton Cerf founded the Internet Society (ISOC) in 1992 “to promote the open development, evolution, and use of the internet for the benefit of all people throughout the world”. One of its first projects was Babel, created in 1997 with Alis Technologies (a language services company) to promote the use of several languages on the web.

---

#### **[pre-web] 1977:**

**The International Federation of Library Associations (IFLA) created UNIMARC as a common bibliographic format for computerised library catalogues. [#4]**

The International Federation of Library Associations (IFLA) created the MARC format (MARC: Machine Readable Cataloguing) in the early 1970s as “a short and convenient term for assigning labels to each part of a catalogue record so that it can be handled by computers”. A few years later, the 20+ national MARC formats in existence meant compatibility issues and extensive editing when bibliographic records were exchanged.

UNIMARC (Universal Machine Readable Cataloguing) was published in 1977 as a solution for cataloguers to be able to process records created in any MARC format. Records in one MARC format could be converted into UNIMARC before being converted into another MARC format. UNIMARC could also be used as a format on its own, as recommended by the European

Commission in 1996. The British Library (with UKMARC), the Library of Congress (with USMARC) and the National Library of Canada (with CAN/MARC) published their own common format MARC 21 in 1999.

---

**[pre-web] 1980:**

**The Pan American Health Organisation (PAHO) began to develop its in-house machine translation software. [#5]**

Developed by PAHO (Pan American Health Organisation), whose main working languages are English and Spanish, the machine translation software SPANAM (Spanish to English) was available in 1980, followed by ENGSPAN (English to Spanish) in 1985. In-house and freelance translators could post-edit the raw output to produce quality translations with a 30-50 percent gain in productivity.

In the 1990s, the software was available on the Local Area Network (LAN) and in field offices, and licensed to public and non-profit institutions in the United States, Latin America and Spain. It was later renamed PAHOMTS, with the addition of language pairs including Portuguese, another official language: English to Portuguese and Portuguese to English in July 2003, Spanish to Portuguese in March 2004, and Portuguese to Spanish in June 2005.

---

**[pre-web] October 1985:**

**Richard Stallman created the Free Software Foundation (FSF) and pioneered the concept of copyleft to ensure the right to use, modify and distribute free software. [#6]**

Richard Stallman, a programmer at the Massachusetts Institute of Technology (MIT), launched the GNU Project in September 1983 as a free software operating system, and founded the Free Software Foundation (FSF) in October 1985 to promote the copyleft and its corresponding licence, the GNU General Public Licence (GPL).

“Copyleft is a general method for making a program (or other work) free (in the sense of freedom, not “zero price”), and requiring all modified and extended versions of the program to be free as well. (...) Copyleft says that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it. Copyleft guarantees that every user has freedom. (...) Copyleft is a way of using the copyright on the program. It doesn’t mean abandoning the copyright; in fact, doing so would make copyleft impossible. The word ‘left’ in ‘copyleft’ is not a reference to the verb ‘to leave’ — only to the direction which is the inverse of ‘right’.” — Excerpt from the GNU Project’s website.

The GNU Free Documentation Licence (GFDL) allows users to copy, redistribute and modify any documentation (manual, textbook, reference or instructional material), commercially or non commercially. The GFDL was the first licence used by Wikipedia when the free collaborative encyclopedia was created in January 2001 – before the Creative Commons licence took over.

---

**[pre-web] 1987:**

**Extended ASCII was a series of text encoding standards including the accented characters of a few languages other than English. [#7]**

The original 7-bit ASCII (first published in 1963) could only process the unaccented characters of the English alphabet. With computer technology spreading outside North America, 8-bit variants of ASCII known as extended ASCII provided up to additional 128 characters (meaning 256 characters in total) in order to process the accented characters of other languages using the Latin alphabet.

The International Organisation for Standardisation (ISO) published these variants in 1987 as ISO 8859 (also called ISO Latin), for example ISO 8859-1 (ISO Latin-1) for the French, German and Spanish alphabets. But there were many conversion and data corruption issues, and many writing systems were left aside, hence the creation of Unicode (first published in January 1991) as a text encoding standard which could process text data in any language around the world.

---

**1990:**

**Tim Berners-Lee invented the World Wide Web and gave his invention to the world. [#8]**

A computer scientist at CERN (European Organisation for Nuclear Research) in Geneva, Switzerland, Tim Berners-Lee developed HTML (HyperText Markup Language), HTTP (HyperText Transfer Protocol) and the first web browser in 1989-1990. The World Wide Web (later known as “the web”) radically changed the way people used the internet. Users could now access text documents with a simple click of the mouse.

The first web browser for the general public was Mosaic (November 1993), which was instrumental in popularising the web. Tim Berners-Lee founded the World Wide Web Consortium (W3C) in October 1994 to develop standards and guidelines for the web, including the ones needed for its internationalisation.

While being interviewed by journalist Pierre Ruetschi for an article in the Swiss daily Tribune de Genève (published on 20 December 1997), Tim Berners-Lee explained that he was “pleased with the rich and diverse information” but he would like “the web to be more interactive, and people to be able to create information together” because the web was intended as “a medium for collaboration, a world of knowledge that we share”.

---

**1990:**

**Anthony Rodrigues Aristar and Helen Dry created the LINGUIST List as an electronic mailing list for linguists. [#9]**

The first electronic mailing list for linguists was created in early 1990 by Anthony Rodrigues Aristar, a professor of linguistics at the University of Western Australia, with 60 subscribers. Linguist Helen Dry joined in 1991 as the co-founder and co-moderator. The list moved to Texas A&M University in 1991 and to Eastern Michigan University in 1997 (and to Indiana University in 2014). The LINGUIST List created its own website in 1997, and became a component of the WWW Virtual Library for Linguistics.

“The LINGUIST List has a policy of posting in any language, since it is a list for linguists. However, we discourage posting the same message in several languages, simply because of the burden extra messages put on our editorial staff. (We are not a bounce-back list, but a moderated one. So each message is organised into an issue with like messages by our student editors before it is posted.) Our experience has been that almost everyone chooses to post in English. But we do link to a translation facility that will present our pages in any of five languages. We also try to have at least

one student editor who is genuinely multilingual, so that readers can correspond with us in languages other than English.” — Helen Dry, co-moderator of the LINGUIST list, interviewed in August 1998.

“We are beginning to collect some primary data. For example, we have searchable databases of dissertation abstracts relevant to linguistics, of information on graduate and undergraduate linguistics programs, and of professional information about individual linguists. The dissertation abstracts collection is, to my knowledge, the only freely available electronic compilation in existence.” — Helen Dry, co-moderator of the LINGUIST list, interviewed in July 1999.

---

#### **January 1991:**

**Unicode was published as a universal text encoding standard that could process text data in any language around the world. [#10]**

Unicode was first published in January 1991 as a text encoding standard providing “a unique number for every character, no matter what the platform, no matter what the program, no matter what the language”. Unicode is maintained by the Unicode Consortium, as well as its variants UTF-8, UTF-16 and UTF-32 (UTF: Unicode Transformation Format). It is a component of the specifications of the World Wide Web Consortium (W3C).

“Although still evolving and only just being incorporated into the latest software, this new coding system translates each character into 16 bits. Whereas 8-bit extended ASCII could only handle a maximum of 256 characters, Unicode can handle over 65,000 unique characters and therefore potentially accommodate all of the world’s writing systems on the computer. So now the tools are more or less in place. They are still not perfect, but at last we can surf the web in Chinese, Japanese, Korean, and numerous other languages that don’t use the western alphabet.” — Brian King, director of the WorldWide Language Institute (WWLI), interviewed in September 1998.

---

#### **December 1992:**

**Projekt Runeberg was the first digital library offering free digitised editions of Nordic literary classics in the public domain. [#11]**

A partner of Project Gutenberg, Projekt Runeberg (Project Runeberg in English) began as a student volunteer project to create free electronic editions of classic Nordic-language literary classics in the public domain. It was founded in December 1992 by Lars Aronsson and his colleagues in cooperation with Linköping University, Sweden, and the computer club Lysator. The project was named after the Nordic poet Johan Ludvig Runeberg. Around 200 e-texts were available in 1998, with a list of 6,000 Nordic authors as a tool for further collection development. Latin literary works and English translations of Nordic literary classics were added later on.

---

#### **1993:**

**John Mark Ockerbloom created The Online Books Page to index all the books freely available online. [#12]**

John Mark Ockerbloom, then a graduate student at the School of Computer Science (CS) of Carnegie Mellon University (CMU), created The Online Books Page in 1993 when the web was still in its infancy with Mosaic as its only browser. The index gave access to 7,000 ebooks in 1998.

“I was the original webmaster here at CMU CS, and started our local web in 1993. The local web included pages pointing to various locally developed resources, and originally The Online Books Page was just one of these pages, containing pointers to some books put online by some of the people in our department. After a while, people started asking about books at other sites, and I noticed that a number of sites had books online, and that it would be useful to have some listing of all of them, so that you could go to one place to download or view books from all over the internet. So that is how my index got started. I eventually gave up the webmaster job in 1996, but kept The Online Books Page, since by then I’d gotten very interested in the great potential the internet had for making literature available to a wide audience.” — John Mark Ockerbloom, founder of The Online Books Page, interviewed in September 1998.

John Mark Ockerbloom was hired as a digital library planner and researcher at the University of Pennsylvania in 1999, and moved The Online Books Page to the website of the University of Pennsylvania Libraries. After listing English-language titles, he added works in other languages as well. The index included 12,000 ebooks in 1999, 20,000 ebooks in 2003 (including 4,000 ebooks published by women), 25,000 ebooks in 2006, 30,000 ebooks in 2008 (including 7,000 ebooks from Project Gutenberg), 35,000 ebooks in 2010, and 2 million ebooks in 2015.

---

#### **April 1993:**

**ABU: La Bibliothèque Universelle was the first digital library offering free digitised editions of French-language literary classics in the public domain. [#13]**

Inspired by Project Gutenberg, ABU: La Bibliothèque Universelle (ABU: The Universal Library) was created in 1993 by ABU (Association des Bibliophiles Universels – Association of Universal Bibliophiles) to create free electronic editions of French-language literary classics in the public domain. It offered 223 works by 76 authors in 1998, and 288 works by 101 authors in 2002.

---

#### **June 1993:**

**Adobe created the Portable Document Format (PDF) to easily distribute and view documents on the web. [#14]**

The two corresponding programs were Acrobat Reader (free, to read PDF documents) and Adobe Acrobat (for a fee, to create PDF documents). As the “veteran” format created in June 1993, PDF was perfected over the years as a global standard for distribution and viewing of information.

“It lets you capture and view robust information from any application, on any computer system and share it with anyone around the world. Individuals, businesses, and government agencies everywhere trust and rely on Adobe PDF to communicate their ideas and vision” — Excerpt from the website in 1998.

10 percent of documents available on the internet were PDF files in 2003, and Acrobat Reader and Adobe Acrobat were available in several languages for several platforms. After being a proprietary format, PDF was released as an open standard in July 2008, and published by the International Organisation for Standardisation (ISO) as ISO 32000.

---

#### **September 1993:**

**Project Gutenberg began to release digitised editions of translated works – English**

## **translations of French literary classics. [#15]**

Founded by Michael Hart in July 1971, [Project Gutenberg](#) was the first digital library offering free digitised editions of English-language literary classics in the public domain.

The first English translations of French-language literary classics were three novels by French novelist [Jules Verne](#): (1) “From the Earth to the Moon” (original title: “De la terre à la lune”), [eBook #83](#) released in September 1993, (2) “Around the World in 80 Days” (original title: “Le tour du monde en quatre-vingts jours”), [eBook #103](#) released in January 1994, and (3) “Twenty Thousand Leagues Under the Sea” (original title: “Vingt mille lieues sous les mers”), [eBook #164](#) released in September 1994.

The first digitised editions of literary classics in languages other than English were French, German, Italian and Spanish literary classics. The first digitised editions of works in French were six novels by [Stendhal](#) and two novels by [Jules Verne](#), all released in early 1997.

“I am publishing in one new language per month right now, and will continue as long as possible.”  
— Michael Hart, founder of Project Gutenberg, [interviewed](#) in July 1999.

---

## **November 1993:**

### **Mosaic was the first web browser for the general public. [#16]**

Developed by the National Center for Supercomputing Applications (NSCA) at the University of Illinois, [Mosaic](#) was distributed free of charge in November 1993. It was instrumental in popularising the World Wide Web, invented by computer scientist Tim Berners-Lee three years earlier.

Part of the Mosaic team migrated to the Netscape Communications Corporation in early 1994 in order to develop a new web browser named [Netscape Navigator](#) (October 1994). Other web browsers were [Opera](#) (April 1995) and [Internet Explorer](#) (Microsoft, July 1995).

---

## **1994:**

### **Projekt Gutenberg-DE was the first digital library offering free digitised editions of German literary classics in the public domain. [#17]**

A partner of Project Gutenberg, [Projekt Gutenberg-DE](#) was created in 1994 to offer e-texts for online reading, with one web page for short texts and several web pages (one per chapter) for longer works. The website also provided an alphabetic list of authors and titles, and a short biography and bibliography for each author.

---

## **1994:**

### **Pierre Perroud created the digital library Athena, which offered free digitised editions of books in the public domain in several languages. [#18]**

Hosted by the University of Geneva, Switzerland, [Athena](#) was founded in 1994 by Pierre Perroud, a teacher at College Voltaire in Geneva. The bilingual French-English website offered electronic editions of works in several languages on various subjects (philosophy, science, literature, history,

economics). Athena offered its own digitised works (200 titles in December 1997) and provided links to works digitised elsewhere (3,500 titles in December 1997, and 8,000 titles one year later).

A dedicated web page offered an index of French-language literary works digitised by Pierre Perroud and others, Geneva being located in French-speaking Switzerland. The web page Helvetia offered documents about Switzerland. The web page Athena Literature Resources offered links to other digital libraries around the world. The website also included the first mineral database available on the web.

---

**1994:**

**“The internet has removed one barrier (distance) but with that comes the barrier of language.” [#19]**

“An interesting thing happened earlier in the history of the internet and I think I learned something from it. In 1994, I was working for a college and trying to install a software package on a particular type of computer. I located a person who was working on the same problem and we began exchanging email. Suddenly, it hit me... the software was written only 30 miles away but I was getting help from a person half way around the world. Distance and geography no longer mattered! OK, this is great! But what is it leading to? I am only able to communicate in English but, fortunately, the other person could use English as well as German which was his mother tongue. The internet has removed one barrier (distance) but with that comes the barrier of language.” — Robert Ware, creator of OneLook Dictionaries, interviewed in September 1998.

---

**May 1994:**

**Tyler Chambers created the Human-Languages Page (H-LP) as the first catalogue of online linguistic resources. [#20]**

“1994 was the year I was really introduced to the web, which was a little while after its christening but long before it was mainstream. That was also the year I began my first multilingual web project [the Human-Languages Page], and there was already a significant number of language-related resources online. This was back before Netscape even existed — Mosaic was almost the only web browser, and web pages were little more than hyperlinked text documents. (...)

“My professional life is currently completely separate from my internet life. Professionally, I am a computer programmer/techie (in Boston, Massachusetts) — I find it challenging and it pays the bills. Online, my work has been making language information available to more people through a couple of my web-based projects [the Human-Languages Page and the Internet Dictionary Project]. While I am not multilingual, nor even bilingual, myself, I see an importance to language and multilingualism that I see in very few other areas. The internet has allowed me to reach millions of people and help them find what they are looking for, something I am glad to do.” — Tyler Chambers, founder of the Human-Languages Page, interviewed in September 1998.

The Human-Languages Page offered a catalogue of 1,800 language-related internet resources in 100 languages in September 1998, with six subject listings (languages and literature, schools and institutions, linguistic resources, products and services, organisations, jobs and internships) and two category listings (dictionaries, language lessons). It merged with the Languages Catalog (part of the WWW Virtual Library) in spring 2001 to become iLoveLanguages, with an index of 2,000 linguistic resources in 100 languages in September 2003.



---

**July 1994:**

**Internet users with a native language other than English reached 5 percent. [#21]**

This number would steadily increase to reach 20 percent in 1998, 50 percent in 2000 and 75 percent in 2015.

---

**October 1994:**

**Tim Berners-Lee, the inventor of the web, founded the World Wide Web Consortium (W3C) to develop standards and guidelines for the web, including for its internationalisation. [#22]**

The World Wide Web Consortium (W3C) soon maintained a web page named “Internationalisation / Localisation” with the protocols needed to create a bilingual or plurilingual website (HTML, base character set, new tags and attributes, HTTP, language negotiation, URLs, and other identifiers including non-ASCII characters). Internationalisation is still part of its mission today, alongside accessibility, privacy and security.

---

**1995:**

**Michael C. Martin created Travlang, the first index of free online translation dictionaries for travellers. [#23]**

Michael Martin, then a physics student in New York, first created the web page “Foreign Languages for Travelers” on the website of his university in 1994. One year later, he moved the web page to his new website Travlang, which quickly became a major portal for both travel and languages. A dedicated web page gave access to free online dictionaries in 16 languages (Afrikaans, Czech, Danish, Dutch, Esperanto, Finnish, French, Frisian, German, Hungarian, Italian, Latin, Norwegian, Portuguese, Spanish, Swedish) in 1998. Another web page offered free online resources to learn 60 different languages. The website also offered links to translation services, language schools, bookshops offering books in several languages, and travel services.

“I think the web is an ideal place to bring different cultures and people together, and that includes being multilingual. Our Travlang site is so popular because of this, and people desire to feel in touch with other parts of the world. The internet is really a great tool for communicating with people you wouldn’t have the opportunity to interact with otherwise. I truly enjoy the global collaboration that has made our ‘Foreign Languages for Travelers’ pages possible.” — Michael Martin, founder of Travlang, interviewed in August 1998.

---

**1995:**

**The WorldWide Language Institute (WWLI, Europe) created NetGlos as a collaborative multilingual glossary of internet terminology. [#24]**

NetGlos – which stands for “Multilingual Glossary of Internet Terminology” – was created in 1995 as an online voluntary, collaborative project for translators around the world. It was available in 13 languages (Chinese, Croatian, Dutch/Flemish, English, French, German, Greek, Hebrew, Italian, Maori, Norwegian, Portuguese, Spanish) in September 1998.

“Much of the technical terminology on the web is still not translated into other languages. And as we found with NetGlos, the translation of these terms is not always a simple process. (...) Our NetGlos project has depended on the goodwill of volunteer translators from Canada, the U.S., Austria, Norway, Belgium, Israel, Portugal, Russia, Greece, Brazil, New Zealand and other countries. I think the hundreds of visitors we get coming to the NetGlos pages every day is an excellent testimony to the success of these types of working relationships. I see the future depending even more on cooperative relationships, although not necessarily on a volunteer basis.” — Brian King, director of the WorldWide Language Institute (WWLI), interviewed in September 1998.

---

**1995:**

**Robert Beard created “A Web of Online Dictionaries” as a directory of free online dictionaries. [#25]**

Created by Robert Beard, a language teacher, in 1995, “A Web of Online Dictionaries” gave access to 800 dictionaries in 150 languages in September 1998 as well as other online linguistic resources (thesauri, vocabularies, glossaries, grammars, textbooks). A web page named “Web of Linguistic Fun” made linguistics appealing for non-specialists. The website was later included in yourDictionary.com, co-founded by Robert Beard in January 2000.

“There was an initial fear that the web posed a threat to multilingualism on the web, since HTML and other programming languages are based on English and since there are simply more websites in English than any other language. However, my website indicates that multilingualism is very much alive and the web may, in fact, serve as a vehicle for preserving many endangered languages. Moreover, the new attention paid by browser developers to the different languages of the world will encourage even more websites in different languages.” — Robert Beard, founder of “A Web of Online Dictionaries”, interviewed in September 1998.

---

**1995:**

**Tyler Chambers created the Internet Dictionary Project (IDP) as a collaborative project to create free online bilingual dictionaries. [#26]**

Tyler Chambers, a computer programmer, first created the Human-Languages Page (H-LP) in May 1994 as an index of free online language-related resources. He launched the Internet Dictionary Project (IDP) in 1995 as a collaborative project to create free online dictionaries from English to six other languages (French, German, Italian, Latin, Portuguese, Spanish).

“Not only is it helpful to the online community to have access to dictionary searches at their fingertips via the World Wide Web, it also sponsors the growth of computer software which can benefit from such dictionaries — from translating programs to spelling checkers to language-education guides and more. By facilitating the creation of these dictionaries online by thousands of anonymous volunteers all over the internet, and by providing the results free-of-charge to anyone, the Internet Dictionary Project hopes to leave its mark on the internet and to inspire others to create projects which will benefit more than a corporation’s gross income. The resulting lists of English words and their translated counterparts are made available through this site to anyone, with no restrictions on their use.” — Excerpt from the website in 1998.

---

**1995:**

**Liber Liber created Progetto Manuzio, the first digital library offering free digitised editions of Italian literary classics in the public domain. [#27]**

Liber Liber, an Italian organisation founded in 1993, describes itself as “promoting artistic and intellectual creative expression, and using computer technology to link the humanities with the sciences”. Progetto Manuzio was its first project, named after the Italian printer and humanist Aldo Manuzio.

“Progetto Manuzio wants to make a noble idea real: the idea of making culture available to everybody. How? By making books, graduation theses, articles, tales or any other document which could be digitised in a computer available all over the world, at any minute and free of charge. Via modem, or using floppy disks (in this case, by adding the cost of a blank disk and postal fees), it is already possible to get hundreds of books. And Progetto Manuzio needs only a few people to make such a masterpiece as Dante Alighieri’s ‘Divina Commedia’ available to millions of people.” — Excerpt from the website.

---

**1995:**

**Major print newspapers and magazines around the world created their own websites to offer an online edition and online archives. [#28]**

Paid electronic editions of a few print newspapers and magazines were available through America OnLine (AOL) or CompuServe in the early 1990s. In the mid-1990s, these newspapers and magazines created their own websites to offer their latest issue in part or in full (often with a free or paid subscription) as well as online archives. Some examples were the daily The New York Times and the daily The Washington Post in the United States; the daily Times and weekly Sunday Times (on their common website Times Online) and the weekly The Economist in the United Kingdom; the weekly Le Monde Diplomatique, the daily Le Monde and the daily Libération in France; the weekly Focus and the weekly Der Spiegel in Germany; and the daily El País in Spain. 3,600 print newspapers around the world offered online editions in December 1997.

---

**1996:**

**The website Merriam-Webster Online provided online editions of print reference works published by Merriam-Webster. [#29]**

Created in 1996, Merriam-Webster Online provided online editions (paid subscription) of the Webster Dictionary, the Webster Thesaurus, the Webster’s Third (a lexical landmark), the Guide to International Business Communications, the Vocabulary Builder (with interactive vocabulary quizzes) and the Barnhart Dictionary Companion (hot new words). Its goal was also to help track down definitions, spellings, pronunciations, synonyms, vocabulary exercises, and other key facts about words and language.

---

**1996:**

**The Ethnologue, a reference catalogue for all living languages, created its own website to offer a free online edition of the catalogue. [#30]**

Published in print by [SIL International](#) every four years since 1971, “Ethnologue: Languages of the World” was a catalogue of minority languages before becoming a reference catalogue of all living languages. A website created in 1996 offered a free [online edition](#) of the Ethnologue as a complement to the print and CD-ROM editions for sale. The online edition offered a full description of the 6,700 living languages spoken in 228 countries.

“The Ethnologue is a catalogue of the languages of the world, with information about where they are spoken, an estimate of the number of speakers, what language family they are in, alternate names, names of dialects, other socio-linguistic and demographic information, dates of published Bibles, a name index, a language family index, and language maps.” — Barbara Grimes, editor of the Ethnologue, [interviewed](#) in January 2000.

The Ethnologue identified 6,909 living languages in 2009, and 7,102 living languages in 2015. A paid subscription model for extensive users in high-income countries was introduced in 2016 in order to sustain the Ethnologue project. Both editions (print and online) are regularly updated with the help of thousands of linguists from partner organisations around the world.

---

**April 1996:**

**Robert Ware created the website OneLook Dictionaries as a “fast finder” in hundreds of free online dictionaries. [#31]**

Robert Ware, a software developer, created [OneLook Dictionaries](#) in April 1996 as a “fast finder” in hundreds of free online dictionaries. It could search among 2 million words from 425 dictionaries covering several subjects (general, business, computer / internet, medical, religion, science, sports, technology, slang) in September 1998.

“I was almost entirely in contact with people who spoke one language and did not have much incentive to expand language abilities. Being in contact with the entire world has a way of changing that. And changing it for the better! I have been slow to start including non-English dictionaries (partly because I am monolingual). But you will now find a few included.” — Robert Ware, founder of OneLook Dictionaries, [interviewed](#) in September 1998.

OneLook Dictionaries could search among 2.5 million words from 530 dictionaries in 2000, 5 million words from 910 dictionaries in 2003, and 19 million words from 1,060 dictionaries in 2010.

---

**April 1996:**

**Brewster Kahle founded the Internet Archive to archive the web for present and future generations before expanding it as a public digital library. [#32]**

[Brewster Kahle](#), a digital librarian and a computer engineer, founded the [Internet Archive](#) in April 1996 as a nonprofit organisation building an “internet library” to offer permanent access to an archive of the web stored every two months, as well as collections of archived web pages on selected topics since 2000.

The [Wayback Machine](#) was launched in October 2001 to crawl the 30 billion web pages archived every two months since 1996, for users to be able to surf the archive of the web by date. It gave access to 65 billion web pages (from 50 million websites) in 2006, 85 billion web pages in 2008, and 150 billion web pages in March 2010.

The Internet Archive also became a digital library of texts, movies, music and software provided by partner organisations. Unlike Google Books, which included copyrighted books and could be accessed from Google's own search engine, the Internet Archive only offered free digitised editions of books in the public domain (except when the copyright holder expressly gave permission) and it could be accessed from any search engine. The Text Archive included 100,000 ebooks in December 2006, 200,000 ebooks in May 2007, one million ebooks in December 2008, and two million ebooks in March 2010.

---

**1997:**

**The first blog was created as an online diary. [#33]**

A blog is an online diary kept by an individual or a group, often in reverse chronological order, and regularly updated (every five minutes or once a month). There were 14 million blogs in July 2005, with 80,000 new blogs per day. Technorati, created as a search engine for blogs, gave the number of 65 million blogs in December 2006, with 175,000 new blogs per day. Some blogs specialise in photos (photo blogs), music (audio blogs or podcasts) and videos (video blogs or vlogs).

Kevin Scannell, a software developer, created the website Indigenous Blogs in September 2011 to offer an index of blogs in indigenous and minority languages, for people to connect and use their languages online.

---

**1997:**

**French publisher Hachette offered a free online edition of its print dictionary of the French language. [#34]**

French publisher Hachette partnered in 1997 with AUPELF-UREF – later known as the Association of Francophone Universities (Agence Universitaire de la Francophonie) – to offer a free online edition of its print dictionary, published in 1995.

The “Dictionnaire Universel Francophone En Ligne” (Online French-Language Universal Dictionary) included standard French as well as the French-language words and phrases used around the world. French was an official language for around 500 million people in 50 countries in 1997.

---

**1997:**

**Some international governmental organisations offered free web versions of their in-house terminology databases. [#35]**

These terminology databases (termbases) were for example:

# ILOTERM, a quadrilingual (English, French, German, Spanish) terminology database maintained by the International Labour Organisation (ILO);

# TERMITE (Telecommunication Terminology Database), a quadrilingual (English, French, Spanish, Russian) terminology database maintained by the International Telecommunication Union (ITU);

# WHOTERM (WHO Terminology Information System), a trilingual (English, French, Spanish) terminology database maintained by the World Health Organisation (WHO).

---

**1997:**

**Eurodicautom, the multilingual terminology database of the European Commission, was freely available on the web. [#36]**

Eurodicautom was initially developed to assist in-house translators working in the Translation Service of the European Commission before being freely available on the web in 1997. It offered economic, scientific, technical and legal terms and phrases in 12 languages – the 11 official languages of the European Union (Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish) and Latin.

There were 120,000 queries per day in late 2003 when Eurodicautom announced a temporary hiatus to prepare for a terminology database in 24 languages with the upcoming enlargement of the European Union. The European Union went from 12 to 25 country members in May 2004, and 27 country members in January 2007. Named InterActive Terminology for Europe (IATE), the new terminology database was freely available on the web in July 2007.

---

**January 1997:**

**The International Labour Organisation (ILO) organised its first symposium on “multimedia convergence”. [#37]**

“Multimedia convergence” could be defined as several information-based industries (printing, publishing, graphic design, media, sound recording, film making) converging into one industry, with information as a common product. A massive loss of jobs in the mid-1990s was a serious enough issue to be tackled by the International Labour Organisation (ILO) in a symposium organised in January 1997 in Geneva, Switzerland. The goal of the symposium was to gather employers, unionists and government representatives from around the world in order “to stimulate reflection on the policies and approaches most apt to prepare our societies and especially our work forces for the turbulent transition towards an information economy”.

Some participants, mostly employers, explained that the information society generated or would generate jobs. Other participants, mostly unionists, explained that there was a rise in unemployment worldwide that needed to be addressed shortly through investment, innovation, vocational training, computer literacy, retraining in digital technology, and fair labour rights, including for teleworkers.

“Some fear a future in which individuals will be forced to struggle for survival in an electronic jungle. And the survival mechanisms which have been developed in recent decades, such as relatively stable employment relations, collective agreements, employee representation, employer-provided job training, and jointly funded social security schemes, may be sorely tested in a world where work crosses borders at the speed of light.” — Footnote in the proceedings of the symposium, 1997.

---

**January 1997:**

**Gabriel was created as the common website of several European national libraries. [#38]**

Gabriel — which stands for “Gateway and Bridge to Europe’s National Libraries” — was a trilingual (English, French, German) website created in January 1997 by the Conference of European National Librarians (CENL). One year later, Gabriel provided links to the internet services

of 38 participating national libraries (Albania, Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russia, San Marino, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, Vatican City).

The internet services of these libraries were OPACs (OPAC: Online Public Access Catalogue), national bibliographies, national union catalogues, and indexes for periodicals. Common projects were indexed on a dedicated web page. Gabriel merged with the website of the European Library (created by CENL in January 2004) in summer 2005 to offer a web portal for 43 national libraries. Europeana was launched in November 2008 as the European public digital library.

---

**April 1997:**

**There were one million websites worldwide. [#39]**

---

**June 1997:**

**Published online by Babel, “The Web Languages Hit Parade” was the first major study on the distribution of languages on the web. [#40]**

A joint project of the Internet Society (ISOC) and Alis Technologies (a language services company), Babel offered a website in seven languages (English, French, German, Italian, Portuguese, Spanish, Swedish) in early 1997, with a typographical and linguistic glossary for each language. The subtitle of the home page was “Towards communicating on [the] internet in any language”. A dedicated web page named “The Internet and Multilingualism” explained how to create a multilingual website.

Babel ran the first major study on the distribution of languages on the web. The results were published in June 1997 on a web page named “The Web Languages Hit Parade” available in seven languages. English (82.3 percent) was followed by German (4 percent), Japanese (1.6 percent), French (1.5 percent), Spanish (1.1 percent), Swedish (1.1 percent) and Italian (1 percent).

---

**June 1997:**

**Caoimhín Ó Donnaíle created a bilingual Gaelic-English index of online resources in all European minority languages on the website of Sabhal Mòr Ostaig. [#41]**

Sabhal Mòr Ostaig is a public higher education college on the island of Skye, Scotland. Caoimhín Ó Donnaíle, a computer teacher, created the bilingual Gaelic-English website of the college in June 1997. The website quickly became the main source of information worldwide on Gaelic, considered an endangered language. He also maintained a dedicated web page offering a bilingual Gaelic-English index of online resources in all European minority languages.

“Students do everything by computer, use Gaelic spell checking, a Gaelic online terminology database [Stòr-dàta]. There are more hits on our website. There is more use of sound. Gaelic radio (both Scottish and Irish) is now available continuously worldwide via the internet. A major project has been the translation of the Opera web browser into Gaelic [by freelance translator Michael Bauer] – the first software of this size available in Gaelic. (...)”

“As regards the future of endangered languages, the internet speeds everything up. If people don’t care about preserving languages, the internet and accompanying globalisation will greatly speed their demise. If people do care about preserving them, the internet will be a tremendous help.” — Caoimhín P. Ó Donnáile, webmaster at Sabhal Mòr Ostaig, interviewed in January 2000.

---

**August 1997:**

**O’Reilly Japan published “The Multilingual Web Guide” (Japanese edition) by Yoshi Mikami, the first book on such a subject. [#42]**

“The Multilingual Web Guide” (Japanese edition) was translated into English (“The Multilingual Web Page Guide”), into French (“Pour un web multilingue”) and into German (“Der Führer zu mehrsprachigen Web-Seiten”) in 1998.

“My native tongue is Japanese. Because I had my graduate education in the U.S. and worked in the computer business, I became bilingual in Japanese and American English. I was always interested in languages and different cultures, so I learned some Russian, French and Chinese along the way. In late 1995, I created on the web [the page] ‘The Languages of the World by Computers and the Internet’ [also known as Logos Home Page or Kotoba Home Page] and tried to summarise there the brief history, linguistic and phonetic features, writing system and computer processing aspects for each of the six major languages of the world, in English and Japanese. As I gained more experience, I invited my two associates [Kenji Sekine and Nobutoshi Kohara] to help me write a book on viewing, understanding and creating multilingual web pages, which was published in August 1997 as ‘The Multilingual Web Guide’ in a Japanese edition, the world’s first book on such a subject.” — Yoshi Mikami, author of “The Multilingual Web Guide”, interviewed in December 1998.

---

**December 1997:**

**3,600 print newspapers around the world offered online editions on the web. [#43]**

The mid-1990s saw the first online editions of some major print newspapers and magazines on the web. Many more newspapers created their online editions during the following months.

“Rapid growth, primarily in Canada, the United Kingdom, Norway, Brazil and Germany, has pushed the total number of non-U.S. online newspapers to 1,563. (...) Elsewhere in America, Mexico has 51 online newspapers, 23 newspapers are online in Central America and 36 are online in the Caribbean. Europe is the next most wired continent for newspapers, with 728 online newspaper sites. After the United Kingdom, Norway has the next most – 53 – and Germany has 43. Asia (led by India) has 223 online newspapers. South America (led by Bolivia) has 161 and Africa (led by South Africa) has 53. Australia and other islands have 64 online newspapers.” — Excerpt from an article by Eric K. Meyer, journalist for AJR/NewsLink, published online in December 1997.

---

**December 1997:**

**Logos, an Italian translation company, made all its translation tools available on the web for free. [#44]**

“We wanted all our translators to have access to the same translation tools [dictionaries, glossaries, grammars, conjugators]. So we made them available on the internet, and while we were at it we decided to make the site open to the public. This made us extremely popular, and also gave us a lot



of exposure. This move has in fact attracted many customers, and it has allowed us to widen our network of translators in the wake of this initiative.” — Rodrigo Vergara, founder of Logos, interviewed by journalist Annie Kahn for her article “Les mots pour le dire” (The words to tell it) published in the French daily Le Monde on 7 December 1997.

The translation tools available online in December 1997 were: (1) the Logos Dictionary, a multilingual dictionary with 7.5 billion words (in September 1998); (2) the Logos Wordtheque, with 328 million words extracted from translated novels, technical manuals and other texts, that could be searched by language, word, author or title; (3) Logos Linguistic Resources, with 500 glossaries; (4) the Logos Universal Conjugator, with verbs in 17 languages.

The Logos Library (formerly Wordtheque) included 710 billion words in 2007, while Linguistic Resources (no change of name) included 1,215 glossaries and Conjugation of Verbs (formerly Universal Conjugator, and later on Logos Conjugator) included verbs in 36 languages.

---

#### **December 1997:**

**Yahoo offered its home page in seven languages for a growing number of internet users with a native language other than English. [#45]**

In late 1997, Yahoo, then a directory of websites, offered its home page not only in English but also in French, German, Japanese, Korean, Norwegian and Swedish. Yahoo was created three years earlier by David Filo and Jerry Lang, two students at Stanford University, California, as an online directory of web sites maintained by human beings (with 63 topics in 1997), while these tasks were fully automated in the search engine AltaVista. However, when a search didn't give any results in Yahoo, it was automatically shunted to AltaVista, and vice versa.

---

#### **December 1997:**

**The search engine AltaVista offered a free instant online translation service in six languages named Babel Fish, developed by SYSTRAN. [#46]**

Originally offered by the search engine AltaVista before moving to the search engine Yahoo in May 2008, Babel Fish could translate a short web page from English into five other languages (French, German, Italian, Portuguese, Spanish), and vice versa. The original web page and the instant translation were face-to-face on the screen. Babel Fish could also translate any short text by copying and pasting it into the web interface. Babel Fish was developed by SYSTRAN (acronym of System Translation), a company specialising in automated language solutions.

“SYSTRAN has partnered with AltaVista and reports between 500,000 and 600,000 visitors a day on [babelfish.altavista.digital.com](http://babelfish.altavista.digital.com), and about 1 million translations per day — ranging from recipes to complete web pages. About 15,000 sites link to Babel Fish, which can translate to and from French, Italian, German, Spanish and Portuguese. The site plans to add Japanese soon.” — Excerpt from journalist Martha L. Stone's article “Web embraces language translation” published online by ZD Network News (ZDNN) on 21 July 1998.

A SYSTRAN-based translation service was also used by Google Translate before it was developed in-house by Google. Other machine translation services were developed by Alis Technologies, Globalink, Lernout & Hauspie and Softissimo, with free and/or paid versions available on the web.

---

**December 1997:**

**There were 70 million internet users — 1.7 percent of the world population. [#47]**

---

**July 1998:**

**Internet users with a native language other than English reached 20 percent. [#48]**

This was a significant milestone. They were only 5 percent in 1994. Internet users with a native language other than English would reach 50 percent in 2000 and 75 percent in 2015.

---

**July 1998:**

**“The number of new non-English websites is expected to outpace the growth of new sites in English.” [#49]**

“This year, the number of new non-English websites is expected to outpace the growth of new sites in English, as the cyber world truly becomes a ‘World Wide Web’. (...) According to Global Reach [a consultancy specialising in the localisation of websites], the fastest growing groups of web newbies are non-English-speaking: Spanish, 22.4 percent; Japanese, 12.3 percent; German, 14 percent; and French, 10 percent. An estimated 55.7 million people access the web whose native language is not English. (...) Only 6 percent of the world population speaks English as a native language (16 percent speak Spanish), while about 80 percent of all web pages are in English.” — Excerpt from journalist Martha L. Stone’s article “Web embraces language translation” published online by ZD Network News (ZDNN) on 21 July 1998.

---

**September 1998:**

**“The internet promotes both a common language \*and\* multilingualism.” [#50]**

“It seems that the internet is moving people in two quite different directions at the same time. The internet (initially based on English) is connecting people all around the world. This is further promoting a common language for people to use for communication. But it is also creating contact between people of different languages and creates a greater interest in multilingualism. A common language is great but in no way replaces this need. So the internet promotes both a common language \*and\* multilingualism. The good news is that it helps provide solutions. The increased interest and need is creating incentives for people around the world to create improved language courses and other assistance and the internet is providing fast and inexpensive opportunities to make them available.” — Robert Ware, creator of OneLook Dictionaries, interviewed in September 1998.

---

**September 1998:**

**“85 percent of the content of the web is in English and going down.” [#51]**

“Because the internet has no national boundaries, the organisation of users is bound by other criteria driven by the medium itself. In terms of multilingualism, you have virtual communities, for example, of what I call ‘language nations’ — all those people on the internet wherever they may be, for whom a given language is their native language. Thus, the Spanish-language nation includes not only Spanish and Latin American users, but millions of Hispanic users in the U.S., as well Spanish-speaking Morocco. (...) 85 percent of the content of the web is in English and going down. This

trend is driven not only by more websites and users in non-English-speaking countries, but by increasing localisation of company and organisation sites, and increasing use of machine translation to/from various languages to translate websites.” — Randy Hobler, marketing consultant for machine translation software and services, interviewed in September 1998.

---

#### **September 1998:**

**The rise of electronic commerce became a major force to promote multilingualism on the web. [#52]**

“Although a multilingual web may be desirable on moral and ethical grounds, such high ideals are not enough to make it other than a reality on a small scale. As well as the appropriate technology being available, there is the impact of electronic commerce as a major force that may make multilingualism the most natural path for cyberspace. Sellers of products and services in the virtual global marketplace into which the internet is developing must be prepared to deal with a virtual world that is just as multilingual as the physical world. If they want to be successful, they had better make sure they are speaking the languages of their customers!” — Brian King, director of the WorldWide Language Institute (WWLI, Europe), interviewed in September 1998.

---

#### **December 1998:**

**1,000 public libraries from 26 European countries created their own websites, according to a report published by the European Commission. [#53]**

These websites ranged from one web page giving the physical address and opening hours of the library to full websites with access to OPACs (OPAC: Online Public Access Catalogue) and other services. The leading countries were Finland (247 libraries), Sweden (132 libraries), the United Kingdom (112 libraries), Denmark (107 libraries), Germany (102 libraries), Netherlands (72 libraries), Lithuania (51 libraries), Spain (56 libraries) and Norway (45 libraries). Newcomers were the Czech Republic (29 libraries) and Portugal (3 libraries). Russia provided a web page with a list of 26 public reference libraries — Source: “Internet and the Library Sphere: Further Progress for European Libraries”, report published online by the European Commission in December 1998.

---

#### **1999:**

**Yves Champollion created Wordfast, a platform-independent translation memory software with terminology processing in real time. [#54]**

Created by Yves Champollion, a translator and software developer, Wordfast was compatible with the IBM WebSphere Translation Server and with CAT tools (CAT: Computer-Assisted Translation) like SDL Trados. A basic version of Wordfast was freely available for a few years, with a manual in 16 languages. In 2010, Wordfast was the top translation memory software on Windows and Mac platforms, and the second one on a Windows platform (after SDL Trados), with more than 20,000 customers, including the United Nations, Nomura Securities, the NASA (National Aeronautics and Space Administration) and McGraw-Hill.

---

#### **January 1999:**

**“The answer is not to ‘fight’ English, much less whine about it, but to build more websites in**

**other languages.” [#55]**

“We can see multilingualism on the internet as a happy and irreversible inevitability. So we have to laugh at the doomsayers who only complain about the supremacy of English. Such supremacy is not wrong in itself, because it is mainly based on statistics (more PCs per inhabitant, more people speaking English, etc.). The answer is not to ‘fight’ English, much less whine about it, but to build more websites in other languages. As a translation service, we also recommend that websites be multilingual. The increasing number of languages on the internet is inevitable and can only boost multicultural exchanges.” — Marcel Grangier, head of the French Section of the Swiss Federal Government’s Central Linguistic Services, interviewed in January 1999.

---

**August 1999:**

**“Worldwide information, but no broad choice for languages, this would be quite a contradiction, wouldn’t it?” [#56]**

“It is very important to be able to communicate in various languages. I would even say this is mandatory, because the information given on the internet is meant for the whole world, so why wouldn’t we get this information in our language or in the language we wish? Worldwide information, but no broad choice for languages, this would be quite a contradiction, wouldn’t it?” — Maria Victoria Marinetti, chemical engineer, translator and language teacher, interviewed in August 1999.

---

**October 1999:**

**Michael Kellogg created WordReference.com to offer free online bilingual dictionaries as well as forums for language professionals. [#57]**

“I started this site in 1999 in an effort to provide free online bilingual dictionaries and tools to the world. The site has grown gradually ever since to become one of the most-used online dictionaries, and the top online dictionary for its language pairs of English-Spanish, English-French, English-Italian, Spanish-French, and Spanish-Portuguese. It is consistently ranked in the top 500 most-visited websites in the world and is a top 100 website in Spain, France, Italy and all of Latin America. I am proud of my history of innovation with dictionaries on the internet. Many of the features that you see in many online dictionaries such as forums and being able to click any word in a dictionary entry were first implemented by me.” — Michael Kellogg, founder of WordReference.com, excerpt from the About web page.

In 2010, WordReference offered dictionaries from English to ten other languages (Arabic, Chinese, Czech, Greek, Japanese, Korean, Polish, Portuguese, Romanian, Turkish), and vice versa, as well as a Spanish-French dictionary and a Spanish-Portuguese dictionary. Conjugation tables were available for French, Italian and Spanish. WordReference Mini, a miniature version of the site, could be embedded into other sites. A mobile version was available for dictionaries from English to French, Italian and Spanish, and vice versa.

---

**December 1999:**

**The print Encyclopaedia Britannica created the website Britannica.com as the online edition of the encyclopedia. [#58]**

[Britannica.com](http://Britannica.com) was the digital equivalent of the 15th print edition of Encyclopaedia Britannica (32 volumes). The website was first available for free, as a complement to the print and CD-ROM editions for sale. It offered links to a selection of articles from 70 magazines, a selection of books and a guide to the best websites, with a single search engine. In July 2001, the website was available for a monthly or yearly subscription fee. In 2009, the encyclopedia welcomed external contributors for its online edition, with registration required to write or edit articles.

---

#### **December 1999:**

##### **Two major French-language print encyclopedias created their online editions. [#59]**

The first French-language encyclopedia available online was [Encyclopaedia Universalis](#), the French-language sister of Encyclopaedia Britannica. It included 28,000 articles by 4,000 contributors, with an annual subscription fee and a number of articles available for free.

The second French-language encyclopedia available online — named Webencyclo (now defunct) — was based on a print encyclopedia published by Éditions Atlas. Its free content was searchable by keyword, topic and media (maps, links, photos, illustrations). A call for papers invited external contributors to submit their articles using the web page Webencyclo Contributif. A free registration was required later on.

---

#### **2000:**

##### **The first wiki was created as a collaborative website. [#60]**

Deriving from the Hawaiian word “wiki” (“fast”), a [wiki](#) is a website allowing several users to collaborate online on the same project by drafting, editing, improving and updating content. It is used for example for dictionaries, encyclopedias and other reference tools. The software can be simple or more elaborate. A simple program handles text and hyperlinks. With a more elaborate program, one can embed images, charts and tables. The most famous wiki is [Wikipedia](#), created in January 2001.

---

#### **February 2000:**

##### **Robert Beard co-founded yourDictionary.com as a web portal for free online dictionaries in every language. [#61]**

Robert Beard, a language teacher, first created [A Web of Online Dictionaries](#) in 1995 before co-founding yourDictionary.com in February 2000 to index free online dictionaries in every language, including in endangered languages.

“Our new website has an index of 1,200+ dictionaries in more than 200 languages.

yourDictionary.com will hopefully be the premier language portal and the largest language resource site on the web. It is now actively acquiring dictionaries and grammars of all languages with a particular focus on endangered languages. It is overseen by a blue ribbon panel of linguistic experts from all over the world. (...)

“Languages that are endangered are primarily languages without writing systems at all (only 1/3 of the world’s 6,000+ languages have writing systems). I do not see the web contributing to the loss of language identity and still suspect it may, in the long run, contribute to strengthening it. More and

more Native Americans, for example, are contacting linguists, asking them to write grammars of their language and help them put up dictionaries. For these people, the web is an affordable boon for cultural expression.” — Robert Beard, co-founder of yourDictionary.com, interviewed in January 2000.

yourDictionary.com offered links to 1,800 dictionaries in 250 languages in September 2003, and to 2,500 dictionaries in 300 languages in April 2007, as well as many other linguistic resources (thesauri, vocabularies, glossaries, grammars, language textbooks).

---

**March 2000:**

**Oxford University Press created the online edition of its print Oxford English Dictionary. [#62]**

Oxford University Press (OUP) created the online edition of the 20-volume Oxford English Dictionary (OED) in March 2000, with a paid subscription and a quarterly update of around 1,000 new or revised entries. Two years later, Oxford University Press (OUP) created Oxford Reference Online (ORO), a comprehensive encyclopedia designed directly for the web, with a paid subscription. According to the publisher, its 60,000 web pages with one million entries were the equivalent of 100 print encyclopedias.

---

**March 2000:**

**There were 300 million internet users — 5 percent of the world population. [#63]**

---

**July 2000:**

**Internet users with a native language other than English reached 50 percent. [#64]**

50 percent was a significant milestone. They were only 5 percent in 1994 and 20 percent in 1998. According to Global Reach, a consultancy specialising in the localisation of websites, they were 52.5 percent in summer 2001, 57 percent in December 2001, 59.8 percent in April 2002, 64.4 percent in September 2003 (including 34.9 percent European internet users and 29.4 percent Asian internet users), and 64.2 percent in March 2004 (including 37.9 percent European internet users and 33 percent Asian internet users). Internet users with a native language other than English would reach 75 percent in 2015.

---

**September 2000:**

**The Quebec Office of the French Language created a major free online French-English terminology dictionary. [#65]**

Designed directly for the web by the Quebec Office of the French Language (Office Québécois de la Langue Française – OQLF), with a database created and maintained by Semantix, the Grand Dictionnaire Terminologique (GDT – Large Terminology Dictionary) offered 3 million terms relating to industry, science and commerce. There were 1.3 million queries during the first month and peaks of 60,000 visits per day. The database was then maintained by Convera Canada, with 3.5 million queries per month in February 2003. A new version of the GDT was available in March 2003, with the database maintained by OQLF itself and the addition of Latin as a third language.

---

**January 2001:**

**Lawrence Lessig founded Creative Commons to adapt copyright to the web and make it easier for creators to share their work. [#66]**

Founded in January 2001 by Lawrence Lessig, a legal scholar, “Creative Commons is a nonprofit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. We provide free licences and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof.”

The first licences were published in December 2002. They were used for example by O'Reilly Media for its books, Wikipedia for its articles, and the Public Library of Science (PLOS) for its academic journals. There were one million Creative Commons licensed works in 2003, 4.7 million works in 2004, 20 million works in 2005, 50 million works in 2006, 90 million works in 2007, 130 million works in 2008, 400 million works in 2010, and 882 million works in 2014.

---

**January 2001:**

**Jimmy Wales and Larry Sanger created Wikipedia as a free collaborative online encyclopedia. [#67]**

Co-founded in January 2001 by Jimmy Wales, an internet entrepreneur, Wikipedia quickly grew into the largest reference website, financed by donations, with no advertising. Its multilingual content is free and written collaboratively by volunteers worldwide, who contribute under a pseudonym. Its website is a wiki, which means that anyone can write, edit, correct and improve the articles. All the articles could be freely used according to the GNU Free Documentation Licence (GFDL), later replaced by a Creative Commons licence.

Wikipedia offered 1.3 million articles in 100 languages in December 2004, 7 million articles in 192 languages in May 2007, 14 million articles in 260 languages in September 2010, and 17 million articles in 270 languages in January 2011 (its tenth anniversary), with 400 million queries per month for all the projects of the Wikimedia Foundation (Wikipedia, Wiktionary, Wikibooks, Wikiquote, Wikisource, Wikipedia Commons, Wikispecies, Wikinews, Wikiversity).

---

**March 2001:**

**The IBM WebSphere Translation Server was a high-end professional product for machine translation in eight languages. [#68]**

IBM embarked on a growing machine translation market with the WebSphere Translation Server, a high-end professional product that could instantly translate web pages, emails and chats from English into eight languages (Chinese, English, French, German, Italian, Japanese, Korean, Spanish), and vice versa, with 500 words processed per second. Other machine translation products and services were offered by SYSTRAN, Alis Technologies, Globalink, Lernout & Hauspie and Softissimo.

---

**February 2002:**

**The Budapest Open Access Initiative (BOAI) was signed as the founding text of the open**

## **access movement for free access to research literature. [#69]**

“By ‘open access’ to [research] literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.” — Excerpt from the [Budapest Open Access Initiative](#) (BOAI).

---

**2003:**

## **The Mobipocket Reader, a popular reading software for ebooks, was available in five languages. [#70]**

Created by [Mobipocket](#), a company specialising in ebooks for mobile devices, the Mobipocket Reader (and its format .mobi) could be used on any PDA (from March 2000), on any computer (from April 2002) and on any [smartphone](#) (from spring 2003). The Mobipocket Reader was available in five languages (English, French, German, Spanish, Italian) in spring 2003, with 6,000 ebooks on the website of Mobipocket and in partner online bookstores. Mobipocket (format, software, ebooks) was bought by Amazon in April 2005 before the launch of its own ebook reader, the Kindle, in November 2007.

---

**2005:**

## **The web celebrated its 15th anniversary. [#71]**

“Less than half of the web is commercial, and the other half is run by passion.” – Wired, August 2005

“The three powers of the internet — ubiquity, variety and interactivity — make its potential use quasi infinite.” – Le Monde, 19 August 2005

One year earlier, the term [Web 2.0](#) was coined by [Tim O’Reilly](#), author, publisher and founder of O’Reilly Media, as the title for a conference he was organising, to describe a web based on community and sharing, with [blogs](#), [wikis](#) and [social media](#). Blogs and social media became major tools for the revival of indigenous and minority languages.

---

**December 2005:**

## **There were one billion internet users — 15.7 percent of the world population. [#72]**

**2006:**

## **An increasing number of mobile phones named smartphones offered access to the web on the go. [#73]**



There were 90 million smartphones (9 per cent) for one billion mobile phones in 2006. The first smartphone was the Nokia 9210, launched as early as 2001, followed by Nokia Series 60 and Sony Ericsson P800. Smartphones replaced PDAs (personal digital assistants), including the veteran models Palm Pilot (launched by Palm in March 1996) and Pocket PC (launched by Microsoft in April 2000).

---

**August 2006:**

**OCLC (Online Computer Library Center) created WorldCat.org to offer a free online version of its union catalogue WorldCat. [#74]**

Created by OCLC, a global library organisation with a website in six languages (English, Chinese, French, German, Portuguese, Spanish) in 1998, the union catalogue WorldCat provided 61 million bibliographic records in 400 languages from 9,000 member libraries in 112 countries in 2005, with a paid subscription. On the new website WorldCat.org launched in August 2006, member libraries could offer free access to 73 million bibliographic records linking to one billion documents, with free or paid access to their digital resources (books, audiobooks, abstracts and full-text articles, photos, music, videos). 1.5 billion documents could be located and/or accessed on WorldCat.org in April 2010.

---

**November 2006:**

**There were 100 million websites. [#75]**

---

**February 2007:**

**ISO 639-3 was published as a set of three-letter codes to identify 7,589 languages (living and extinct, ancient and reconstructed, major and minor, written and unwritten), including sign languages. [#76]**

The new ISO standard was compiled in 2004-2006 by SIL International, the publisher of the Ethnologue, a reference catalogue of all living languages. Published by the International Organisation for Standardisation (ISO) in February 2007, the standard ISO 639-3 (2007) reconciled the set of three-letter identifiers used in the Ethnologue since the inception of its database in 1971 with the set of 400 three-letter codes used in the previous standard ISO 639-2 (1998), as well as the set of three-letter codes developed by the LINGUIST List for ancient and constructed languages. The first ISO standard was ISO 639-1 (1988) as a set of two-letter language codes. SIL International was named the registration authority for the language identifier inventory, and has administered the annual cycle of changes and updates since then.

---

**May 2007:**

**The Encyclopedia of Life was created as a global scientific effort to document all known species of animals and plants. [#77]**

The aim of the Encyclopedia of Life was not only to document all known species of animals and plants (1.9 million), including endangered species, but also to expedite the millions of species yet to be discovered and catalogued (about 8 million). A single web portal would give access to a wealth

of documents (texts, photos, maps, audio files, videos) scattered online and offline, with a web page for each species.

This project was first described by American biologist and naturalist [Edward Wilson](#) in an essay dated 2002. Technology improvements made it possible five years later, with content aggregators, mash-up, wikis and large-scale content management to process multimedia files. The first pages of the encyclopedia were available in mid-2008. The English edition was expected to be translated into several languages by partner organisations.

---

#### **June 2007:**

**InterActive Terminology for Europe (IATE) was freely available on the web as the new multilingual terminology database of the European Union. [#78]**

The first multilingual terminology database of the European Union was [Eurodicautom](#), available on the web in 12 languages from 1997 to 2003. Eurodicautom was replaced by [IATE](#) (InterActive Terminology for Europe), a larger terminology database in 24 languages as an answer to the enlargement of the European Union. The European Union went from 12 to 25 country members in May 2004 (IATE began its intranet service at that time), and 27 country members in January 2007. IATE became a free public service on the web in June 2007.

Maintained in Luxembourg by the Translation Centre of the European Union institutions, IATE included 1.4 million entries in the 23 official languages of the European Union (Bulgarian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, Swedish) and Latin in June 2007. It offered 8.4 million words, 540,000 abbreviations and 130,000 phrases in December 2009.

---

#### **October 2007:**

**After using a SYSTRAN-based translation service, Google launched Google Translate as its own free online translation service. [#79]**

The SYSTRAN-based translation service (first used in [Babel Fish](#)) was available in 13 languages (Arabic, Chinese-simplified, Chinese-traditional, Dutch, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish), with several stages for language pairs. To get an instant translation, users could either copy and paste a text in the web interface or supply a hyperlink. [Google Translate](#) was developed in-house from October 2007. As stated at the time, “Google Translate can help users understand the general content of a foreign language text, but doesn’t deliver accurate translations.”

A speech program to read the translated text was available in 2009. Different translations for the same word were available in 2011. Released in June 2009 (and shut down in December 2019), the [Google Translator Toolkit](#) was a free web service for (human) translators to edit the translations generated by Google Translate, first with English as a source language and 47 target languages, and then with more language pairs.

---

#### **January 2008:**

**Unicode superseded ASCII (American Standard Code for Information Interchange) as the**

## **main text encoding standard on the web. [#80]**

Unicode (first published in January 1991) took over 7-bit ASCII (first published in 1963) and 8-bit extended ASCII (first published in 1987) as the main text encoding standard on the web.

---

## **November 2008:**

### **Europeana was created as the European public digital library. [#81]**

Europeana offered two million documents when it went live in November 2008, 6 million documents in March 2010, and 10 million documents on a revamped website in September 2010. The previous European project was the European Library, which was not a digital library but a web portal for 43 national libraries, created in January 2004 by the Conference of European National Librarians (CENL). The first web portal created by CENL, named Gabriel, was available from 1997 to 2005 before merging with the European Library.

---

## **2010:**

### **The “Atlas of the World’s Languages in Danger” published by UNESCO offered a free online edition alongside its 3rd print edition for sale. [#82]**

UNESCO (United Nations Educational, Scientific and Cultural Organisation) published the free online edition of its atlas in three languages (English, French, Spanish) alongside its print edition (3rd edition, 2010) for sale. The 2,473 languages could be searched by country and area, language name (including alternate names), number of speakers from/to, language vitality, and ISO 639-3 code. The alternate names were spelling variants, dialects and names in non-Roman scripts.

---

## **October 2010:**

### **Project Gutenberg offered free digitised editions of literary classics in the public domain in 59 languages and in several formats. [#83]**

While most ebooks were in English in the 1990s, with a few literary classics in French, German, Italian and Spanish, Project Gutenberg included ebooks in 25 languages in early 2004, ebooks in 42 languages (including Sanskrit and the Mayan languages) in July 2005, and ebooks in 59 languages in October 2010.

The ten main languages on 7 October 2010 were English (28,441 ebooks), French (1,659 ebooks), German (709 ebooks), Finnish (536 ebooks), Dutch (496 ebooks), Portuguese (473 ebooks), Chinese (405 ebooks), Spanish (295 ebooks), Italian (250 ebooks) and Greek (101 ebooks). These languages were followed by Latin, Esperanto, Swedish and Tagalog.

Many literary classics were digitised and proofread by Distributed Proofreaders, created in 2000 by Charles Franks to support the digitisation of books in the public domain. These ebooks (20,000 ebooks as of April 2011) were sent as “unique titles to the bookshelves of Project Gutenberg, free to enjoy for everybody. Distributed Proofreaders is a truly international community. People from all over the world contribute.”

---

**March 2011:**

**There were 2 billion internet users — 30.2 percent of the world population. [#84]**

---

**March 2011:**

**Kevin Scannell created the website Indigenous Tweets to index tweets in indigenous and minority languages, for people to connect and use their languages online. [#85]**

Kevin Scannell, a software developer, first designed a statistical software crawling the web to find Twitter threads before indexing them on Indigenous Tweets. He identified tweets in 35 languages in March 2011, tweets in 71 languages in April 2011, tweets in 144 languages in March 2013, and tweets in 184 languages in October 2017.

“Together we’re breaking down the idea that only global languages like English and French have a place online! The primary aim of Indigenous Tweets is to help build online language communities through Twitter. We hope that the site makes it easier for speakers of indigenous and minority languages to find each other in the vast sea of English, French, Spanish, and other global languages that dominate Twitter. Even speakers of languages like Basque and Welsh with vibrant online communities have been surprised to find just how many people there are tweeting in their language. This is the other goal of Indigenous Tweets: it’s a message to the world that says ‘We are here and we’re proud of our languages’. For languages with just a few users, I hope it inspires some people to start — make your voice heard!” — Kevin Scannell, creator of Indigenous Tweets, excerpt from his blog.

---

**September 2011:**

**Kevin Scannell created the website Indigenous Blogs to index blogs in indigenous and minority languages, for people to connect and use their languages online. [#86]**

After creating Indigenous Tweets in March 2011, Kevin Scannell, a software developer, created a similar website for blogs, Indigenous Blogs, in September 2011. He began with the blogs hosted by Blogspot (which also hosts his own blog), WordPress and Tumblr. He identified blogs in 50 languages in September 2011, blogs in 74 languages in March 2013, and blogs in 85 languages in October 2017.

“Speakers of indigenous and minority languages around the world are struggling to keep their languages and cultures alive. More and more language groups are turning to the web as a tool for language revitalisation, and as a result there are now thousands of people blogging and using social media sites like Facebook and Twitter in their native language. These sites have allowed sometimes scattered communities to connect and use their languages online in a natural way. Social media has also been important in engaging young people, who are the most important demographic in language revitalisation efforts.” – Kevin Scannell, creator of Indigenous Tweets and Indigenous Blogs, excerpt from his blog.

---

**March 2015:**

**There were 3 billion internet users — 42.3 percent of the world population. [#87]**

---

**May 2015:**

**There were one billion websites. [#88]**

---

**July 2015:**

**Internet users with a native language other than English reached 75 percent. [#89]**

75 percent was a significant milestone. Previous milestones were 50 percent in 2000, 20 percent in 1998 and 5 percent in 1994.

---

**October 2015:**

**“A central storage space for translations would be useful for localisation projects in endangered languages.” [#90]**

Michael Bauer, a freelance translator from English to Gaelic, considered an endangered language, has been a driving force behind many Gaelic online resources. His software localisation projects with fellow localiser GunChleoc (a pseudonym) included the Gaelic versions of the web browser Opera (as early as 2001), Google Chrome, OpenOffice, LibreOffice and Accentuate.us, and the Gaelic language packs for Microsoft Windows and Microsoft Office.

In his spare time, Michael Bauer worked for ten years to offer a digital edition of Dwelly, the main Gaelic-English dictionary, before teaming up with software developer Will Robertson to maintain Am Faclair Beag, a bidirectional online dictionary incorporating Dwelly with more modern content.

“There are, sadly, far too few users and there are some aspects which actually actively limit usage. For example, Gaelic schools cannot install Gaelic software because the IT contracts are given by the councils to outside IT companies who only provide English software and operating systems. Because they limit the admin rights of the users at schools, this means it is very difficult to install software which is not on their official ‘list’ and because Gaelic is not mentioned in the contract, they don’t put it there. Free and open software has helped carve out more of a space on the web for Gaelic, and cooperating with commercial long-term partners is helping to produce some very useful enabling technologies. (...)”

“A central storage space for translations would be useful for localisation projects, with a shared translation memory, thus avoiding to endlessly re-translate the same terms, phrases and sentence segments. If the translations could be available from the same site, like a meta-Poodle [a community localisation server], everyone working for the revival of a minority language on the web would benefit from it. (...) Part of the problem is there are so many platforms these days, all trying to carve out a niche.” — Michael Bauer, translator from English to Gaelic, interviewed in October 2015.

---

Copyright © 2024-25 Marie Lebert

Licence CC BY-NC-SA version 4.0

Written by marielebert

2024-10-15 at 09:33

Posted in Uncategorized